JANUARY 1910 $3.00 a Year
50c Per Copy

The Journal of Educational Psychology

Editors
J. C. BAGLEY, C. E. SEASHORE,
J. CARLETON BELL, G. M. WHIPPLE.

PUBLISHED MONTHLY, FROM JULY TO JUNE, BY
WARWICK & YORK, Inc.,
HERSHEY, PA. BALTIMORE, MD.
[Entered as second-class matter at the postoffice at Hershey, Pa., under the Act of March 3, 1879.]

CONTENTS

ARTHUR W. SCHELLE. *The Importance of Diagnosis in Educational Measurement* 1
W. H. UHL. *Monetary Fees for College Professors* 13
LEWIS M. TERMAN. *Some Data on the Binet Test of Naming Words* 29
HANOR A. WEBB. *A Preliminary Test in Chemistry* 36

COMMUNICATIONS AND DISCUSSIONS

The Effect of the Summer Vacation on Ability in the Fundamentals of Arithmetic. (Maurice A. Garfinkel) 44
A Comparison of the Results of Group Tests and the Point Scale Test. (W. B. Pyle and G. E. Bronce) 48

ABSTRACTS AND REVIEWS

E. A. DOLL. *Clinical Studies in Feeble-Mindedness.* (J. M. McCallie) 51
A. J. ROSANOFF, HELEN MARTIN AND ISABEL R. ROSANOFF. *A Higher Scale of Mental Measurement and its Application to Cases of Insanity.* (G. M. Whipple) 53
EDITORIAL 54
NOTES AND NEWS 56
PUBLICATIONS RECEIVED 58

Lecture Charts and Slides

Mounted and Ready for Teaching Agriculture in Schools

SIMPLE, PRACTICAL, IMPRESSIVE

Successful Teaching of Agriculture Assured. Send 10 cts. in postage for sample of booklet

Agricultural Extension Department

INTERNATIONAL HARVESTER COMPANY of N. J.

Harvester Building, Chicago

Price revision now effective on certain

Books and Magazines

Published by

WARWICK & YORK, Inc.

BALTIMORE, MD.

(Postage 5c per book East of the Mississippi. 8c per book West of the Mississippi and to foreign countries.)

A POINT SCALE FOR MEASURING MENTAL ABILITY. Yerkes-Bridges-Hardwick. 216 pages. $1.60.

BALTIMORE COUNTY (MARYLAND) PUBLIC SCHOOLS COURSE OF STUDY. GRADES I-VIII. Tall-Davidson-Cook. 700 pages. $2.75.

CHILDREN'S PERCEPTIONS. Winch. 240 pages. 5 color frontispiece. $1.60.

HOW I KEPT MY BABY WELL. Noyes. 193 pages. 30 illustrations. $1.40.

MEMORY AND THE LEARNING PROCESS. Lyon. 184 pages. 9 folded plates. $2.00.

THE EXPERIMENTAL DETERMINATION OF MENTAL DISCIPLINE IN SCHOOL STUDIES. Rugg. 150 pages. 3 plates. $1.50.

THE MENTALITY OF THE CRIMINAL WOMAN. Weidensall. 350 pages. illus. $2.00.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY. Single issues, 50 cents. Back volumes $6.00. Subscriptions paid in advance remain at $3.00 per year.

EDUCATIONAL ADMINISTRATION AND SUPERVISION. Single issues 40 cents each. Back volumes $4.00. Subscriptions paid in advance remain $2.50 per year.

WARWICK & YORK, Inc.

Educational Publishers

BALTIMORE

OUR LATEST PUBLICATION

Work-While Stories for Every Day

By LAYTON B. Evans

Price, $1.50

If you tell stories, you need this book

It provides in one volume a story for every day of the school year, without repetition. No other story book embodies this feature.

Send for descriptive circular

MILTON BRADLEY COMPANY

S. E. Cor. 17th & Arch Sts., Philadelphia, Pa.

For Mental Testing

SYLLABUS FOR THE CLINICAL EXAMINATION OF CHILDREN

Edmund S. Huey

Teachers, social workers, physicians, and all who have to make mental estimates of individuals, will appreciate the separate and revised publication in a convenient booklet, of Dr. Huey's syllabus for the Clinical Examination of Children. The Syllabus points out and classifies the significant facts to be obtained as to the child's heredity, environment, and personal history of growth, disease and accident. It provides for a teacher's report of the child's life and progress in school, for a physical and medical examination, and for a measurement of mental age. The varied blank tests are given in a form most intelligible for record, and there are explicit directions for giving the Binet tests. The original Binet pictures have been reproduced, thus permitting the picture tests to be given as standardized. The Syllabus is published in paper binding so as to reduce the selling price to a minimum, thus enabling those training social workers and teachers to use the booklet with students who might not be able to procure a copy of a more costly book.

24 Pages, 25 Cents. $20.00 Per Dozen.

BINET-SIMON TEST FOR INTELLIGENCE AGE

Prepared by L. B. Hoag

Shortly before his death, Dr. Huey prepared for the press all clinical work at the Johns Hopkins Hospital, a convenient record blank (4 pp., 6¼x10, which makes intelligible and yet concise statement of the results of Binet examinations — "Binet-Simon Test for Intelligence Age." Uniform with the "Binet-Simon Tests for Intelligence Age" are four other blanks for use in the examination of defective and backward children. These supplementary blanks are "Home Record and History," "School Report," "A Point Scale of Tests for Intelligence," and "Directions for Giving and Scoring the Tests." A 3x5 clinical reference card sums up, for filing, the main results of the entire examination and study of the child. These blanks will greatly expedite and simplify the actual work of examining children.

Specimen set of six forms, 25 Cents.

Duplicate copies of any form, 40 Cents.
$1.25 per Hundred Copies.

WARWICK & YORK, Inc.

Educational Publishers

BALTIMORE, MD.

VOLUME X JANUARY, 1919 NUMBER 1

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

THE IMPORTANCE OF DIAGNOSIS IN EDUCATIONAL MEASUREMENT

ARTHUR W. KALLOM

Department of Educational Investigation and Measurement, Boston, Massachusetts

THE GROWTH OF EDUCATIONAL MEASUREMENT

Since the movement for educational measurement first started, there has been a great increase in the literature and in the number of tests which are available for measuring educational results in the elementary schools. The movement began through the demand for the educational survey. This survey was made by experts in education from outside the city system. It was the intention to thus obtain an impartial and just statement of the conditions as they existed in the city which was being surveyed. The movement has spread until by 1917 the United States Commissioner of Education has reported over ninety different surveys. Also there are a large number of cities which employ their own director,[1] whose work consists in giving standard tests from time to time and measuring the results of the educational product in the city in the same unprejudiced and impartial manner which is obtained by having the outside expert come in to do the testing, with the advantage that the director becomes familiar with the city, its civic and educational policies, and may thus make his plans fit in with other movements.

The school surveys have shown to the educational profession the lack of three important elements in our educational administration.[2]

(a) Lack of standards of achievement
(b) Lack of standard tests
(c) Lack of a record of educational experience

Lack of any standard achievement is shown by the uncertainty in the mind of the teacher of what ought to be expected from any pupil in the school system. That is, how many examples of a cer-

[1]Educational Standards, vol. V, May, 1915, Boston.
Ballon, F. W., [2]School and Society, vol. V, No. 108, pages 61 to 70.

tain kind should be performed in addition within a given time? What per cent. of accuracy should be expected of a pupil in the seventh grade in subtracting examples of a certain kind within a definite period of time? How much of the geography which was taught in the fifth grade should be retained by pupils in the eighth grade? The answers to all of these questions and others like them are very indefinite in the teacher's mind.

This lack of any standards of achievement is directly dependent upon the lack of standard tests. The ordinary examination given by a teacher is not a standard test because the teacher does not know the elements of which the test consists. That is, she does not know the relative value of the different questions of which the test is composed. Even if the test partakes of purely mechanical matter as the adding of columns of figures, the teacher does not know whether all combinations have been included, nor does she take into consideration the relative difficulty of the different combinations. So it has been only since the inception of educational measurement, that standard tests have been available and even now tests along certain lines are difficult to prepare.

The lack of any record of educational experience is nearly, if not quite, as great as the lack in the other two respects. Many of our good teachers, however, have kept a record of their experience in various subjects from time to time. For example, some teachers have kept a list of the words which have been missed by children in their composition work and have used these words as a basis of their spelling lessons. They have also kept more or less of a record in connection with their arithmetic. These, however, have been incomplete and have not been systematized. Neither has this material been organized and passed on so that others may profit by the mistakes which have been made. In general the more complex a subject or topic, the less detailed is the record which has been kept.

Important Phases of Educational Measurement

The inauguration of any department of educational measurement in a public school system is only defensible on the grounds that it has for its ultimate end the better education of the individual pupil. All departments of the school administration including superintendent, supervisors, principals, and teachers are for the sole pur-

pose of educating the pupil in the most efficient manner possible. The maintenance of any one of these departments can only be justified if it performs this particular service.

Standards of Achievement

How may educational measurement help the teacher in performing this service of educating the individual boy or girl? The first work of any department in educational investigation is to set up a definite standard of achievement in the different branches of the curriculum as fast as possible. This standard should be based not on any theoretical basis, but upon the actual work of the pupils taking a standard test under supervised conditions. It is not possible to discuss in this paper what those conditions may be. Suffice it to say at this time, that those conditions must be such that the person in charge of the test knows what they are.

Record Cards Showing Achievement

This standard now becomes a goal which may guide the teacher in the instruction of the children. Her effort and energy may now be directed toward those pupils who have not reached the standard of achievement as set by the city instead of having her work distributed over a class, *part of which does not need the extra work which is being given.* These individuals are pointed out because the teacher keeps a record of each test on a card which is uniform for the city and is sent with the pupils as they go from class to class in the same school or from school to school within the city.

After this standard of achievement has been found, the pupil also has a knowledge of what he is expected to do. He knows whether his achievement is above or below the standard set for the city. If he is below, he becomes anxious immediately to raise his attainment and come nearer to the standard as set for the city. If he is above the standard, he then immediately becomes anxious to maintain his record. Thus the setting of a standard of achievement on a city-wide basis becomes a very important motive in his work.

Figures 1, 2, and 3 show the records of three individuals in work with Courtis tests, series B, with the explanation, diagnosis, and treatment in each case. The solid line in each figure indicates the standard, while the dotted line indicates the score of the individual.[1]

[1]Taken from School Document No. 22, Boston Public Schools, "*The Value to the Teacher, to the Principal, and to the Superintendent of Individual and Class Records from Standard Tests.*"

RECORD OF ARITHMETIC TESTS

Name A. — M. — Girl (Boy or Girl) Age at first trial 13 (On last birthday)

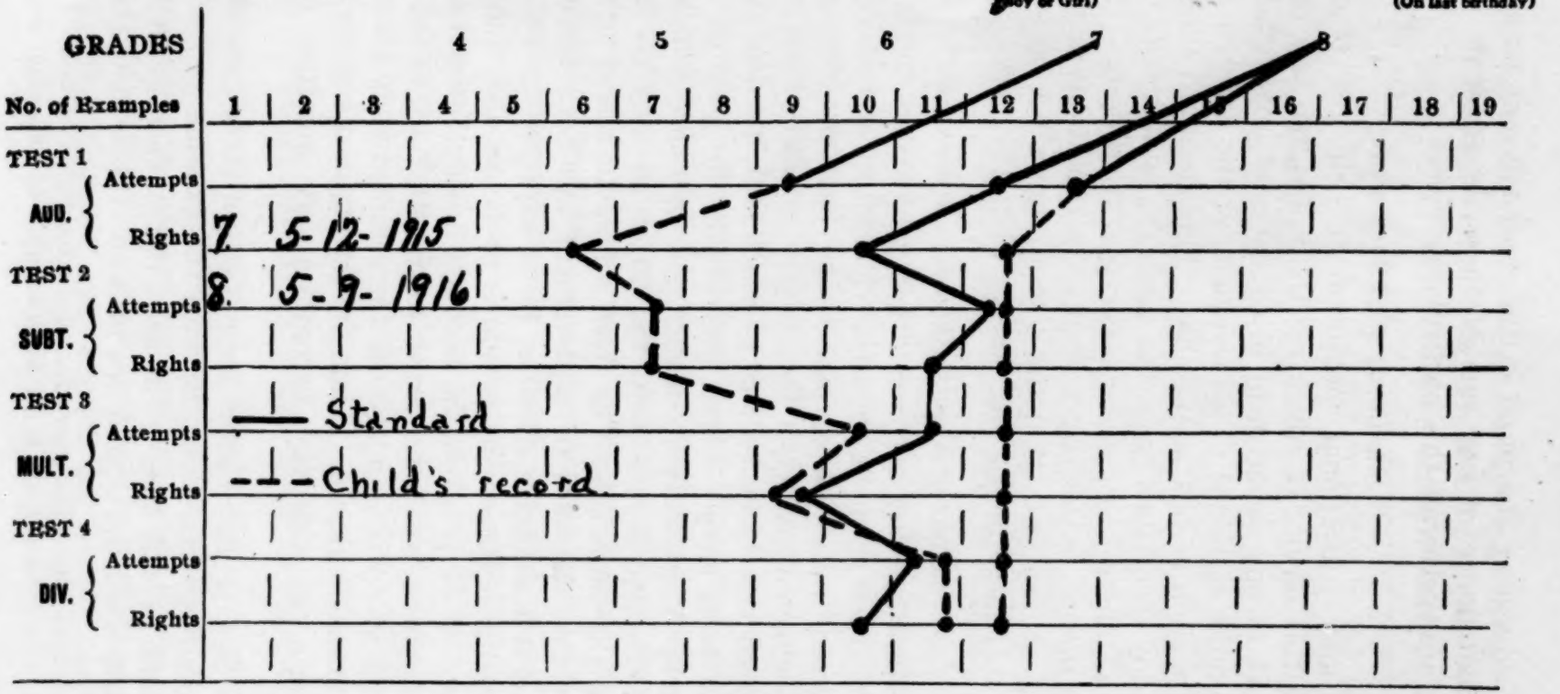

FIG. 1

Interpretation of Fig. 1

EXPLANATION.—This shows the graph of a girl graduated from an elementary school in June, 1916, who was thirteen years of age when she first took the test. Her record shows two tests: May 12, 1915, when she was in Grade VII, and May 9, 1916, when she was in Grade VIII.

DIAGNOSIS.—When she first took the test she showed uneven ability in the processes. She was slow in addition as well as inaccurate. She was accurate in subtraction but slow. Her multiplication score was up to eighth-grade standard in the number of examples correct, while the division score was up to the eighth-grade standard in speed and showed perfect accuracy. In the second test she showed remarkable uniformity in the four processes, attempting twelve examples in every process but addition, in which she attempted thirteen, and in every case getting twelve examples correct.

TREATMENT.—The second score is an unusual one and shows an ideal toward which to work. While the number of examples attempted is at or near the standard, there is at the same time almost perfect accuracy. Just enough emphasis has been placed on the deficiencies in addition and subtraction to eradicate them and bring ability in those processes up to the ability in multiplication and division. The record of this girl is the kind of standard to which the department looks forward. It would constitute a better standard than the present one. When pupils enough to make a change reasonable have succeeded in achieving a result such as this girl's, the present standard will be modified accordingly.

RECORD OF ARITHMETIC TESTS

Name A ——— B ———, Girl (Boy or Girl) Age at first trial 11 (On last birthday)

School District ..... School ..... City Boston

DATE	GRADE		SERIES B							
			TEST 1 ADD.		TEST 2 SUBT.		TEST 3 MULT.		TEST 4 DIV.	
			Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights
1-8-1914	VI	1st Trial	7	2	5	1	5	4	4	0
4-21-1914	VI	2nd Trial	8	5	5	1	5	3	3	0
10-9-1914	VII	3rd Trial	9	5	5	2	8	6	4	2
5-17-1915	VII	4th Trial	12	12	11	2	11	11	6	0
5-11-1916	VIII	5th Trial	17	17	20	14	14	12	19	19
		6th Trial								
		7th Trial								

FIG. 2

Interpretation of Fig. 2

EXPLANATION.—A——B——graduated from an elementary school in May, 1916. She took the test five times; on January 1, 1914, and April 21, 1914, when she was in the sixth grade, on October 9, 1914, and May 17, 1915, when she was in the seventh grade, and on May 11, 1916, when she was in the eighth grade.

DIAGNOSIS.—The addition score improves at every trial. From the score of 7–2 it finally goes to 17–17, perfect accuracy. In subtraction the increase in the number of examples attempted is just as marked; the growth in accuracy is not so noticeable, although it is evident that the accuracy rises from 20 per cent (5 attempted, 1 right) to 70 per cent (20 attempted, 14 right). The multiplication score, which shows good accuracy in the first test, although slow in attempts, rises to 14 examples attempted with 12 right. The division score in the first and second tests shows that the girl attempts only 3 or 4 examples and is unable to get 1 right. In the third test she finds herself able to perform correctly 2 examples out of a total of 4 attempted. In the next test she attempts more, but lapses to none right. Between that and the fifth test she evidently learns how to divide, as shown by her next score of 19–19.

TREATMENT.—This girl was exceptionally inaccurate in every process except multiplication. Drill has been applied so that noteworthy progress has been made in those processes which needed it most. In addition and division progress is especially noticeable. The deficiencies have been eradicated. There has been perhaps a little over-emphasis as the scores exceed the standard. For example, in 1915 the addition and multiplication scores had reached the eigh-grade standard. Therefore, this child needed only enough drill on those processes to keep the record at the standard.

RECORD OF ARITHMETIC TESTS.

Name J—— L——, Boy (Boy or Girl) Age at first trial 13 (On last birthday)

School District ..... School ..... City Boston

DATE	GRADE		SERIES B							
			TEST 1 ADD.		TEST 2 SUBT.		TEST 3 MULT.		TEST 4 DIV.	
			Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights
1914-1-8	7	1st Trial	8	5	9	5	5	4	5	0
1914-4-23	7	2nd Trial	10	2	20	13	13	10	7	3
1915-5-12	7	3rd Trial	14	6	13	12	11	6	8	6
1916-5-9	8	4th Trial	13	7	18	15	14	8	9	2
		5th Trial								
		6th Trial								
		7th Trial								

FIG 3

Interpretation of Fig. 3

EXPLANATION.—This is the record of a boy who graduated at the age of sixteen from an elementary school. He spent two years in Grade VII and three trials of the test are recorded in that grade. The other trial recorded is when he was a pupil in Grade VIII.

DIAGNOSIS.—By an examination of this boy's record it will be seen that in four successive tests he made little or no improvement in his addition and division scores, even though they were shown to be low in the first test. There is variability also in the subtraction and multiplication scores with some improvement in speed.

TREATMENT.—This report has proved of no value to the teacher as she did not use it to diagnose the pupil's difficulty and prescribe for him the particular kind of drill he needed. It illustrates well the futility of class drill in eliminating individual difficulties. The child goes blundering on, meeting the same difficulty and making the same kind of mistake. The record of the first test showed especially that something serious was wrong with this boy's ability in addition and division. Yet his teacher did not make proper use of that record and the inability showed itself again in the next test. His next teacher did not use the record and it went on until this boy graduated from an elementary school with the ability to perform correctly only two long division examples in eight minutes and only seven addition examples in the same length of time.

Use of Record of Arithmetic Tests

Such records give the information necessary to know which pupils need individual help. By means of these cards the teacher is able to determine immediately upon the entrance of a class into her room what pupils are poor in any given branch of the work which has been tested. This not only saves time ordinarily used for finding out these facts, but insures that the knowledge which the teacher has is much more dependable than any she might gather for herself. Not only does she find out from these records that a particular pupil may be poor in fundamentals, but she finds in what particular operation the pupil is poor. For example, the pupil may be relatively poor in arithmetic, but his record may

show that his poor work is determined largely by his work in addition and multiplication, whereas in subtraction and division his work may be exceedingly good.

Urgent Need for Methods of Diagnosis

The knowledge, however, that a particular pupil is weak in addition does not necessarily tell the teacher in what particular phase of addition the individual may be weak. The results obtained in such a standard test as the Courtis standard tests in arithmetic, simply indicate a condition and do not in themselves diagnose this condition. A low score is a symptom; the interpretation of the sympton is quite another matter. For example, there are a number of different factors entering into the addition of integers. A pupil weak in any factor would necessarily show a poor record in adding examples consisting of nine numbers of three digits each. Some of the more important factors in the operation of addition are as follows:

1. Simple addition combinations
2. Single column addition of three figures each
3. "Bridging the tens" as 38 plus 7
4. Column addition, seven figures
5. Carrying
6. Column addition with increased attention span, thirteen figures to the column
7. Addition of numbers consisting of one or more digits

Thus it will be seen that a pupil may be weak in his ability to carry from one column to the next and, therefore, be weak in his ability to do addition examples. On the other hand, he may be able to carry correctly but his span of attention may not be sufficiently strong to enable him to add columns of eight or nine figures and for this reason he would show a poor result in his ability to add integers.

When the teacher has found out the fact that a pupil is weak, it becomes necessary for her to use her resourcefulness to determine at what stage the pupil is weak. This is made comparatively easy by the best forms of practice material which are on the market at the present time.

SUGGESTIONS FOR DIAGNOSIS

What is meant by diagnosis? By diagnosis is meant the taking of certain symptoms that exist and finding out from them what

the trouble is. Such diagnosis would naturally be followed by proper treatment. Each pupil may have a different trouble, and, therefore, require a different treatment. Thus the importance of the individual is brought to the front. There may be two reasons for looking over a set of papers. First, to simply mark a pupil's paper; second, to gain some knowledge which will serve as a basis for specific corrective work. The latter method of correcting work is diagnosis.

There are three principles which underlie any diagnosis of a pupil's paper.

(a) Trouble with the individual pupils cannot be discovered by studying and drilling the children who have examples right. The principle that success comes through failure must be practised here and the teacher will find her richest material for study in the poorest papers.

(b) After selecting the papers with low scores, it is necessary to find out from the paper, or from studying corrective material how the pupil does his work. The fact that most pupils do their work in the same way makes this possible. Thus it will be necessary to work with the individual and so determine the habits he has formed which lead him to do his work in the wrong way.

(c) The most important step of all is the keeping of a record of each pupil's performances. In no other way can one tell whether the pupil is advancing or not.

Conclusions

3. One of the most important phases brought out as a result of giving standard tests has been the emphasis upon and need of a more careful diagnosis of all work in the elementary school. This is possible only when all phases of the topic under consideration are represented. Tests which do not give all the possibilities may point out an important symptom but cannot be used for purposes of diagnosing the trouble of the pupil.

4. After the results of a test have been studied, drill work, adjusted to the nature of the difficulty, should be given. Pupils who are able to do the mechanical work sufficiently well should not be required to take this drill work required of slower pupils. A method

which requires the entire class to do all the drill work is just as inefficient and unpedagogical as is the method which would allow the slow pupil to drift along and do what he could with only class instruction. Any method of diagnosis emphasizes the individual. Depending on averages and citing of averages ignores and covers up the individual.

1. The mechanical parts of the different subjects must be reduced as nearly as possible to automatism.

2. The various mechanical operations in different subjects are not one general habit but made up of a number of specific habits. For example, *there is no one general habit of adding integers but a number of specific habits involving each possible type in this process.*

MENTALITY TESTS FOR COLLEGE FRESHMEN

W. L. UHL

Northwestern University

The aim of this article is to report the results of a series of tests given in the fall of 1916 to a group of one hundred freshmen at Northwestern University. The purpose of giving the tests was to secure norms for freshmen when tested in large groups, in the hope that such norms might be used in making classifications of students.

Descriptions of the Tests Used

a. The Completion Tests. The Trabue Completion Tests K and M were used. By giving two completion tests, it was believed that the results would show the value of the tests for discovering certain abilities of freshmen. Both of the sets which were used are designed for measuring the abilities of persons of high school age or older. The time allowances for K and M were five minutes and three minutes, respectively; otherwise, the standard directions for giving and scoring the tests were followed.

b. The Hard Opposites Tests. For the opposites test a list of twenty words was used. Two minutes were allowed for this test, the students having written their names upon the backs of the test sheets before turning them.

c. The Information Test. A one-hundred-word list was used for this test. About thirty new words were substituted for an equal number of words of Whipple's list. These substitutions were made because too many words of the original list are new to nearly all freshmen. There was also an attempt to represent, by the words, such fields of knowledge as freshmen may reasonably be expected to have explored. An attempt was made to arrange the words so that every fourth word would be a relatively easy one; *e. g.*, base-hit, Anthony Wayne, X-ray, etc., were separated by words which were believed to be of greater difficulty.

Another change in this test has to do with the responses of the subjects. In its original form, the subject was asked to place one of four letters before each word: D, if he could define it as accurately as would the dictionary; E, if he could explain it clearly enough to give one who was not familiar with it some idea of its meaning, although unable to give an exact definition; F, if the word was only roughly familiar, so that he could not use it intelligently;

and N, if the word was entirely new to him. This method admitted of marked over-or under-estimation, by the subject, of his own ability. In order to have a check upon the alleged knowledge of the words marked, high school subjects taking the original Whipple test were asked to define,* on the reverse side of the blank, the first five words marked D and the first five marked E. The following are some of the definitions given:

"Anthony Wayne—an author.

Architrave—an arch formed by trees.

Base-hit—a three base hit meaning the runner hit the ball far enough so he could get back to home before its return.

Caedmon—is one of the cycles of ancient times.

Catalepsy—a disease or sickness of the mind which often causes fits or entire loss of memory.

Chamfer—a kind of medicine.

Clearing-house—a wholesale house that is selling out.

Euclid—a Latin book of translations; a character in Latin; an ancient personage who was a great hero.

Les Miserables—a piece of music."*

With the exceptions of the responses to the words Catalepsy and Clearing-house, it will be seen that there is a clear case of over-estimation which would have been avoided in marking the tests if the subjects had been required to give even brief definitions as responses. In the hope of securing a better type of responses, the following directions were given to the freshmen.

The purpose of this test is to measure very roughly the breadth of your general information. On the sheet of paper which you have, there is a list of words taken from a large number of fields of knowledge. Your part in this test will be to show that you know from what field of knowledge each word is drawn.

(To illustrate the exact method to be used, the following examples were written on the blackboard).

Bryant—literature (poor but better than no response).
—American literature (adequate).

Snowbound—poem
—American poem or Whittier's poem.
(Comments as to responses were made as under Bryant).

Caesar—history.
—Roman History.

punt—sport.
—foot-ball.

Paris—city.
—French city.

C. O. D.—business term (adequate).

dyne—science.
—physics

*JONES, E. E.: An Unpublished Report of a Survey of the Schools of Cambridge, Ill.

Merely state the field of knowledge unless it is easier for you to define the word. Try to use no more than three words for each word of the test—in many cases, one word will be sufficient. Take the words in order, beginning with the first column. Work as rapidly as you can.

This was the last of a series of tests all of which had been timed tests: as a result, the freshmen thought they were to have but a limited amount of time for this. After ten minutes, it was seen that practically all were through work—nearly all had stopped; a very few worked longer. No one worked more than twelve minutes.

Results of the Tests

a. The Completion Tests. As the two completion tests were given less than five minutes apart, it was supposed that, with few exceptions, each student would make about the same relative score in one test as in the other. It seemed plausible, however, that there would be less difficulty in getting started upon Test M than upon K, owing to the practice effect. It is possible also that a period of three minutes was too short for Test M. A comparison of the results indicates, however, that much the same degree of success was attained in one test as in the other. The number receiving the maximum score of two on each sentence was practically the same; this tends to the conclusion that the students were given sufficient time to do as well in one test as in the other. This conclusion is supported by other considerations. The number attaining a score of one on different sentences was small; with the exception of the scores for sentence 5, in Test M, there was a close similarity between the results of the two tests in this respect. The scores for *any* written, though futile, attempt upon a sentence are the same in the two tests except in case of sentences 6–8. The explanation of this fact is probably that an insufficient amount of time was given in Test M for making futile attempts to complete these more difficult sentences; some influence may possibly be due to the students' unwillingness, by the time these sentences were reached, to make attempts which were unlikely to be successful. The latter explanation seems untenable, however, because many were willing to make futile attempts upon sentence 6.

In Table I are shown the frequencies of scores for the two tests. The maximum scores for K and M are 14 and 16, respectively. It will be seen that all lacked two or more of attaining the maximum score in M, while three lacked only one of attaining the maximum score for K. Either the necessary shortening of the time for M or

TABLE I

Results of the Completion Tests

Scores		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	or more Total
Frequencies	K	0	1	2	2	4	3	13	10	16	11	22	0	13	3	0	0	100
	M	0	0	2	0	5	3	9	12	18	11	19	9	6	2	4	0	100

Medians K 8.9
M 9.1

TABLE II

Results of the Opposites Test

Scores	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	Total
Frequencies	1	0	1	6	17	17	26	12	11	7	2	100

Median 31.2

the greater difficulty of the last sentences may account for this difference.* The medians for the two tests indicate that M is the easier test even disregarding the difference in time.

So far the tests seem to have yielded results which are quite similar. When we consider the scores of individual students, however, we see that a very large number shifted their positions greatly. The more striking fact is that practically as many made a low score in the first test as in the second; 42 made a higher score in K than in M; 11 made the same; 47 made more in M.

b. The Opposites Test. The results of the opposites test are shown in Table II. The scheme of scoring suggested by King and Gold† was followed. The essential features of the scheme are as follows: (1) for exact opposites a score of 3 is given; (2) for closely allied but not the best, or for words commonly but loosely used as opposites a score of 2 is given; (3) for words of only remotely opposite signification a score of 1 is given; (4) a score of 0 is given for the following kinds of words: random or coined words, phrases, words of different parts of speech, and synonyms. The responses indicate that sufficient time was given as the last words were tried by nearly all of the students, and practically all worked straight through the list.

---

*The values assigned by Trabue do not support the latter possibility. Pp. 26–7, *loc. cit.*

†IRVING KING AND HUGO GOLD, *A Tentative Standardization of Certain "Opposite Tests."* Journal of Educational Psychology, Vol. VII, pp. 459–482, Oct., 1916.

c. The Information Test. The information test proved to be the most difficult to score. In selecting words for the test there was an attempt (1) to eliminate the possibility of chance correct reactions; (2) to find words of such character that a satisfactory reaction consisting of only one word would be possible; and (3) to avoid words having merely local or transient significance. In scoring, the reactions were classified as follows: (a) correct (score of 2); (b) not wrong but too general, vague, or far-fetched (score of 1); (c) still more slight relation to the stimulus-word (score of ½); (d) wrong, or inadmissible on account of being arbitrarily considered as too far-fetched (score of 0). The distribution of scores is shown in Table III.

TABLE III Information

Distribution of Scores for the ~~Opposites~~ Test

Score	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130 or more	Total
Frequencies	1	2	11	14	11	13	16	14	9	3	5	1	0	100

Median 68.4

In the assignment of values, some cases were treated very arbitrarily, as in the case of "religion" as a reaction to "Zionism;" Zionism is an especially poor word for the test as scored because it admits of a response from all subjects if they are willing to use a very general response. "Religion" as a response to Zionism is, therefore, believed to be much easier than "philosophy" as a response to "Stoicism," although both responses were given equal value. In case of the word "Utopia," a higher value was assigned to the response "ideal land" than to "More's poem," the latter being considered as worthy of a score 1, while "English poem" was given a score of ½. In some cases, as, for example, the last-mentioned response, the less definite and more common response, "English literature," was given a score of 2. This distinction was made because the response, "English literature," fulfills all the requirements of the directions for the test. Some responses were given certain values which result from the fact that their stimulus-words are often found outside their precise fields of knowledge; *e. g.*, "spoils system" may as properly be assigned to American history

as to political science according to the directions for the test; also, *Habeas corpus* is quite as properly an historical as a legal term except that the latter is more definite. If objections be raised in regard to the responses, they should be maintained only after consultation with advanced students from several different departments. It is believed that the judgments of a group of graduate students and instructors will necessitate a considerable number of arbitrary assignments of values, owing to the different relations in which certain words have often been met. It is, of course, obvious that many responses require only the merest smattering of acquaintance with the words of the test. Inasmuch as only proper nouns were capitalized, a penalty was fixed for overlooking capitalization as in the case of Millet which was called "grain."

Discussion of Results

In order to compare the results of the tests with those of university courses, the grades of the freshmen in their first semester courses in required English and mathematics were secured. Also, the three mathematics instructors who had these one-hundred students in mathematics were asked to rank their own students as to general ability. For this ranking, the following simple instructions were given:

> So far as you are acquainted with the following persons, mark down after each name a number indicating the intellectual fitness of each individual. Grade in terms of what you consider to be their real ability or capacity to do difficult intellectual work—not in terms of their customary class performances.
>
> Mark +2 after names of superior individuals;
> Mark +1 after names of very good individuals;
> Mark 0 after names of merely average individuals;
> Mark −1 after names of individuals just below average;
> Mark −2 after names of very poor individuals.

In addition to these checks upon the tests, the university registrar supplied data showing the standing of the freshmen in the high school classes in which they were graduated.

In Table IV are the correlations between the different tests and between the tests and English and mathematics, abilities as judged by instructors, and ranks in high school classes.* The most strik-

*The correlations given in this table were secured by means of the Spearman "footrule for correlation." The values for "R" are here given in terms of values given in Rugg's *Statistical Methods Applied to Education*, page 402.

ing correlation is that between Completion Tests K and M. In case of a relatively homogeneous group of subjects such as university freshmen may be assumed to be, many shiftings of ranks may be expected, but, even so, one would expect a higher correlation in case of similar tests which are considerably different. That is, if the correlation between tests K and M is only .32, it would not be expected that test M and the information test should have a correlation of .42, and then that K and the information test should have a correlation of only .18. The instability of ranks is again

TABLE IV

Correlations between Tests and University Grades

	Compl. K	Compl. M	Oppo-sites	Informa-tion	English	Mathe-matics
Compl. K.....	...	.32	.21	.18	.37	.16
Compl. M....	.32	...	.26	.42	.38	.41
Opposites.....	.21	.26	...	.26	.44	.29
Information...	.18	.42	.26	...	.38	.43
English.....	.37	.38	.44	.38	...	.48
Mathematics..	.16	.41	.29	.43	.48	...

(K and M) and English	.39
(K and M) and Mathematics	.36
Instructors' Judgments and Mathematics	.93
Judgments and High School Ranks	.59
Mathematics and High School Ranks	.55
Mathematics and English	.48
High School Ranks and All Tests	.37
Mathematics and All Tests	.41
English and All Tests	.46
Judgments and All Tests	.36

shown in the case of K and M in their respective correlations of .16 and .41 with mathematics; when the scores for K and M were combined, the correlation with mathematics was .36.

A much higher correlation was expected between the completion tests and English, and, in order to find out whether the low correlation was due to the fact that the tests were as a whole too easy, the scores received by the students upon the last four sentences of K were found. These scores were as follows:

Score	0	19	Score	5	2
	1	4		6	10
	2	18		7	3
	3	14		8	0
	4	30			

This part of the test seems to be too hard as is shown by the bunching of the scores at the lower end. The correlation for this array with English was .29, even lower than that for K in its entirety.

Table V shows how marked were the individual shiftings in the different tests and university courses. The term "quarter" denotes the positions of the *ranks* of students; for example, students whose ranks fell between 1 and 25 were classified as being within the first quarter, between 25 and 50 as the second quarter, etc. This differs from the quartile divisions used in part in the next two tables, because it often happened that the rank of a large number of students was so placed as to throw too many into a given quarter; for example, too many fall into the second quarter for K because a large group received the rank of 25.5. A comparison of various forms of groupings indicated that for our purposes the grouping by quarters is quite as valid and satisfactory as grouping by quartiles. Either form of grouping is arbitrary here, because in several of the tests and in the university courses quartile divisions would not be based upon scores or grades alone, owing to the "crowding" of certain parts of the distribution.*

Table VI shows the degree of accuracy which may be expected in the results of the tests. Division A of this table shows that only a small number of those who received grades of B or C were displaced as much as two quarters; that is, of those who were placed in quartile 3 by Test K, four received a grade of either A or B in English. (It will be remembered that there were only sixteen A's in English). Also, eleven of quartile 1 in K received either C, D, or F in English. These comparisons are made with English because the test is a language test and because the comparisons are still more unfavorable for the test when made with mathematics.

Before comparing the other tests, let us note the relation between mathematics and English, Division D of Table VI. Of those who are in quartile 1 of English, only two are in the third or fourth quarters in mathematics. Now, turning to K and M, we find many more shiftings except for quartile 3. These findings would tend to the conclusion that if one desired to test a student's ability to pursue English successfully, much more precise results would be se-

*There was the following distribution of grades in English: A, 16; B, 27; C, 49; D, 2; F, 6. Quartile I includes nine who received a grade of B.

TABLE V

Relative Positions of Individuals in Different T sts

A. With reference to K.

Quarter in K	Test	Position in different quarters in other tests			
		Quarters in other tests			
	M	1st	2d	3d	4th
1		4	7	2	1
2		10	10	6	7
3		5	10	5	7
4		3	3	7	13
1	Opp.	1	9	2	2
2		9	9	8	7
3		8	7	6	6
4		3	4	7	12
1	Inf.	3	4	3	4
2		10	8	10	5
3		5	5	12	5
4		3	6	7	10
1	Eng.	2	6	6	0
2		6	11	14	2
3		6	3	17	1
4		2	7	12	5
1	Math.	6	4	3	1
2		13	9	5	6
3		10	6	5	6
4		8	10	3	5

B. With reference to English.

Quarter in Eng.	Test	Position in different quarters in other tests			
	M	1st	2d	3d	4th
1		3	8	4	1
2		9	8	6	4
3		9	14	8	18
4		1	1	0	6
1	Opp.	7	4	1	4
2		5	13	5	4
3		7	12	15	15
4		2	1	2	3
1	Inf.	5	5	3	3
2		8	6	6	7
3		10	12	16	11
4		1	0	2	5
1	K	2	6	6	2
2		6	11	3	7
3		6	14	17	12
4		0	2	1	5
1	Math.	7	7	1	1
2		18	8	1	0
3		12	13	12	12
4		0	2	1	5

cured if the student were given a semester's work in mathematics than if he were given either K or M or both. In fact, the shiftings are so great in all the tests that from one-fifth to one-third are displaced as much as would be indicated in terms of university grades by two or three letters; for example, from a grade of A as shown by one test to a grade of C or D as shown by another test. Table VII shows, however, that the method of scoring may have much to do with the shiftings between English and the completion tests.

TABLE VI

This Table Shows the Number of Individuals Who Were Displaced Two or More Quarters by Certain Tests

A. With reference to K.						B. With reference to M.						
Quartile in K	M	Opp.	Inf.	Eng.	Math.	Quartile in M	K	Opp.	Inf.	Eng.	Math.	Judg.
1	3	8	11	11	7	1	9	8	7	11	6	6
2	7	6	5	1	3	2	2	5	7	1	3	2
3	4	7	8	4	9	3	2	4	6	7	10	14
4	5	8	7	9	17	4	8	8	7	3	10	11
Total	19	29	31	25	36		21	25	27	22	29	33

C. With reference to K and M combined.								D. With reference to English.					
Quartile in K–M	K	M	Opp.	Inf.	Eng.	Math.	Judg.	Quartile in Eng.	K	M	Opp.	Inf.	Math.
1	2	0	8	7	10	8	8	1	8	7	7	9	2
2	3	1	4	7	0	2	2	2	14	11	8	7	2
3	0	0	3	4	2	9	9	3	0	4	5	5	9
4	1	1	9	7	8	13	14	4	19	12	9	7	9
Total	6	2	24	25	20	32	33		41	34	29	28	22

In case of the completion tests, grades have been assigned to the respective scores as follows: Scores for the two tests may be from 0 to 30; the scores were, therefore, divided into five equal parts and assigned grades corresponding to their respective positions. This table shows a greater value for the completion tests than is shown by any of the other tables. The correlation (Pearson formula) for mathematics and English is .48, while that for English

TABLE VII

Correlation Charts for English and Mathematics and for English and the Completion Tests

A. English and Mathematics

		Grades in English					
		A	B	C	D	F	Total
Mathematics	A	3	9	0	0	0	12
	B	5	9	11	0	0	25
	C	6	8	17	1	2	34
	D	1	1	10	0	0	12
	F	1	0	11	1	4	17
Totals		16	27	49	2	6	100

B. English and the Completion Tests combined scores of K and M

Scores		Grades in English A	B	C	D	F	Total
24–29 99	A	0	1	1	0	0	2
18–23 99	B	11	15	17	0	2	45
12–17 99	C	5	9	24	0	0	38
6–11 99	D	0	2	7	2	3	14
0– 5 99	F	0	0	0	0	1	1
Totals		16	27	49	2	6	100

and the completion tests with this distribution is only .27. Even with this low correlation, it will be seen from Table VIII that there are fewer displacements from the grades in English in the case of the completion tests than in the case of mathematics. The results as shown in this table indicate that an instructor might, by

TABLE VIII

Displacements from Grades in English

	Grades in English A	B	C	D	F	Total
K and M	5	2	1	0	2	10
Math.	8	1	11	0	2	22

a proper method of adjusting his scoring of both his regular class work and the two completion tests, form a fairly accurate opinion of the abilities of his students from such tests alone. Tables V to VIII also indicate very clearly that coefficients of correlation conceal the very things for which an instructor is looking, that is, individual achievements.

In Table IX are given the correlations between some of the tests as indicated by both Spearman and Pearson formulae. While there are considerable differences in case of several pairs of coefficients, the range for the entire group is about the same.

In Table X are shown the ranks of a few individuals in all tests. Division A shows the ranks in other tests of those students who ranked from 1–5 and from 96–100 in K and M combined; as will

TABLE IX

Correlation Coefficients as Shown by Different Formuale.

	Spearman	Pearson
Completion Tests K and M	.32	.38
Test M and Opposites	.26	.42
Test M and English	.38	.34
Test K and English	.37	.24
Opposites and English	.48	.30
Tests K and M and English	.40	.27*
English and Mathematics	.48	.48

*Based upon Table VII.

be seen, even these extremes are not rigidly placed by the tests. Division D is probably the most valuable part of this table. The students who ranked 1 and 97, respectively, are good examples of those who are likely to be misplaced by brief tests. In individual tests given to them later, No. 1 was alert and rapid except in the Healy Puzzle-box test where both were very slow. No. 97 showed less interest in the tests and worked them more slowly. In regard to the university grades, No. 1 resented the opinion that there might be a close relation between scores made in tests and class grades; she said that she did not like mathematics nor her instructor. She appears to be the type that will respond readily and rapidly to a social situation. On the other hand, No. 97 is far more phlegmatic; also, her instructor ranks her very high in general ability. One of these subjects thus appears to be able in short tests to surpass her long-continued efforts, while the other can do little unless given ample time for reflection.

TABLE X

Ranks in All Tests of Those Receiving Very High or Very Low Scores

DIVISION A

Ranks in K and M	Ranks in other tests				Grades	
	K	M	Opp.	Inf.	Eng.	Math.
1	2	2.5	39	5.5	B	A
2	9	2.5	12.5	12	C	D
3	42	2.5	88	49.5	B	A
4	9	18	43	48	A	F
5	2	32	35.5	16.5	B	B
96	93.5	92.5	2	30	C	D
97	98.5	87	49	54.5	B	A
98	100	76.5	75.5	7	C	C
99	93.5	96	100	98	F	F
100	98.5	96	16.5	66	F	C

DIVISION B

Ranks in Opposites	Ranks in other tests				Grades	
	K	M	K & M	Inf.	Eng.	Math.
1	42	10	17.5	47	C	B
2	93.5	92.5	96	30	C	D
3	25.5	18	17.5	93	B	A
4	55.5	87	77.5	60	C	F
5	55.5	76.5	72	8	B	A
96	25.5	47	32.5	70	C	B
97	25.5	76.5	54	58	C	C
98	81	61.5	71.5	69	C	C
99	25.5	47	32.5	76.5	C	D
100	93.5	96	98	98	F	F

DIVISION C

Ranks in Inf.	Ranks in other tests				Grades	
	K	M	K & M	Opp.	Eng.	Math.
1	55.5	32	42.5	69.5	C	F
2	55.5	87	77.5	7	A	A
3	42	18	24.5	16.5	B	B
4	69	5.5	24.5	69.5	F	C
5	2	2.5	1	39	B	A
96	42	87	72	69.5	C	C
97	25.5	32	24.5	20	F	F
98	93.5	96	98	100	F	F
99	42	32	32.5	20	B	A
100	42	98	87	94	F	F

DIVISION D

Ranks in All Tests	Ranks in each test					Grades	
	K	M	K & M	Opp.	Inf.	Eng.	Math.
1	9	2.5	2	12.5	12	C	E
2	2	2.5	1	39	5.5	B	A
3	25.5	10	11	20	14.5	B	C
4	25.5	18	17.5	25.5	5.5	B	B
5	42	18	24.5	16.5	25	B	B
96	93.5	87	92	75.5	83	C	C
97	81	92.5	87	78	91.5	B	A
98	81	87	85	86.5	88.5	C	F
99	81	96	92	84.5	99	F	F
100	93.5	96	98	100	98	F	F

Summary and Conclusions

1. As is true of tests in general, these tests fail to measure both accurately and adequately. While a test may call for the exercise of a general ability, yet it can call forth only a slight or fragmentary type of reaction; as a consequence, further testing by means of a similar test may give quite different results. This fact is illustrated in case of the completion tests. It seems likely that tests similar to the information test might give very different results.

2. If, however, standardized tests can be made adequate, they will offer a superior means for testing general abilities. An instructor can devise "quizzes" which test both general and special abilities; these are open to the objection that they cannot well be standardized as to either the method of giving or the amount of demand made upon general as distinguished from special preparation. A standardized test should reveal either special strength or weakness; a form of weakness may be that of being unable to do well in a brief intense test, or inability to use language forms accurately and automatically. The completion tests if given upon several different days in several different forms should reveal such weaknesses. By means of such a form of comparison as that shown in Table VII it was seen that there was as much likelihood of arriving at approximately correct judgments of the students tested by noting their achievements in the completion tests as in mathematics.

3. In order to make the results most reliable, individual testing may be preferable, but distributed periods as just suggested seem, from the results of these tests to be satisfactory so far as group work is concerned.

4. The homogeneity of the group upon which this report is based probably has much to do with the failure to secure high correlations. In such a group there is more competition for the same places or ranks than is true in case of a group in which some members are very far below and others far above the median. While there is marked difference between the ability of a student who attains high marks and one who attains low marks, there is a far greater difference between the extremes in the sixth grade of elementary school, that is, before highly selective processes have been long at work.

5. Of the tests given, the information test seems the least practicable. Such a test might be very valuable in giving an index of ability in any given field of knowledge; as a general test, however, its results are vague, hard to evaluate, and show no higher correla-

tion than the other tests. It has, of course, the advantage of indicating something that is not shown by any of the other tests. More careful selection of words from different fields and a standardized method of scoring are necessary if the test is to be of greatest value.

6. In attempting to devise tests and to secure standards which will show a high correlation with school work, one faces the problem of testing in a brief interval of time those special or general capactities that determine one's ability to carry on successfully a very complex type of activity for an extended period of time and under varying conditions. The problem seems to be too complicated to be explained merely upon the basis that the tests measure more highly specialized capacities or different capacities than those found in a university course such as English. For example, the concomitance of abilities to succeed in English and mathematics is more frequent than that of abilities to do well in tests involving the use of English forms and the university course in English. While mere concomitance of abilities may show no necessary relation between them, yet it is, on the other hand, possible that problem-solving of one kind or another enters into the two university subjects sufficiently to make their demands more nearly the same than the demands of English and the tests. The social situations involved in such courses are certainly much more nearly the same than those in tests and regular work.

7. Again, the low correlations may be due in part to the relative simplicity of tests as compared with complex courses. Besides the social situations just mentioned there is in one a long-continued demand upon one's will to preserve in case of a university course, while a test is a novel form of activity which, though often seeming quite impossible, is of very short duration. There must be ability to set one's self promptly at the test in a way which the semester's work may not require; this was shown in case of student Number 97 in Division D of Table X. The effect of seeing one's neighbors finishing a test probably has the effect of interfering greatly with the forms of automatic responses and ability to discriminate as demanded by the test. One of the great difficulties in working through a test is often the fact that one becomes obsessed with an idea in the form of a word or phrase; he then tries to force a solution. In this period of comparative destitution, any form of distraction may cause failure.

8. At the outset of this paper, it was said that one of the aims was to try a method of testing; the results secured seem to warrant the use of the method in case of tests in which a standard amount of time is to be given to all. Monitors are, of course, needed to prevent efforts to work overtime. As to the practicability of the tests, the completion tests, and the opposites seem to be better adapted for such use than the information test on account of reasons stated under "5." It is believed that the instructor would secure valuable information concerning his students' abilities by studying the results of this kind of tests. The medians for the tests are shown in Tables I–III and may be considered as norms. The fact that either weak or strong students may, by the amount of effort which they put forth, secure high marks in university courses, makes impossible an absolute judgment as to one's ability; for this reason, the results of tests may be of value in planning needed assistance to certain students thus aiding in checking the appalling "mortality" of college freshmen.

SOME DATA ON THE BINET TEST OF NAMING WORDS*

LEWIS M. TERMAN

Leland Stanford, Jr. University

Some of the questions which suggest themselves regarding the 60 word test are the following:

1. To what extent does the test correlate with other measures of intelligence?
2. What number of minutes alloted to the test will yield the best measure of intelligence?
3. Does the slope of the curve for words named in the six successive half minutes differ for children of different degrees of brightness?
4. To what extent is the ability to name words in a free association test dependent on size of vocabulary?
5. Do older and younger children of the same mental age differ in the number of words named?
6. What sex difference does the test reveal?
7. What is the significance of such features of the response as stereotypy, improvisation of words, the size of the association groups, the nature of the associations involved, the degree of abstractness of words used, etc.

Questions relating to the psychological analysis of the response are especially interesting and important, but the writer is not prepared to deal with them here. The present paper will be limited to the presentation of certain quantitative data bearing chiefly on the value of the test as a measure of mental ability.

Correlation with Mental Age

Table I shows for a miscellaneous group of 480 children the correlation between mental age and number of words named in three minutes.

The correlation of .535 is considerable, but not as high as might have been expected. Had there been a larger proportion of our subjects of extremely high and also of extremely low mental age, the correlation would have been higher. The large majority of subjects entering into Table I are between 8 and 13 in mental age.

*The writer is indebted to Miss Bess Henry.

TABLE I

Total Words in 3 Minutes

Mental Age	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	Total
17[e]																		1								1	2
17																											0
16[e]																				1	1	2					4
16												1															1
15[e]												1		1								1					3
15													1		1			3									5
14[e]													1		1	1		2									5
14													2	1	1	1											5
13[e]											1		3	3	2		2	1				1					13
13											4	4	2	2	1	1	2	3	2	2					1		24
12[e]						1			2		2	3	4	4	2	2		1	2						1	1	25
12								1	3	2	5	7	4	2	2	3	1		1	1							32
11[e]						3	1		1	1	3		1	1	6	3	1			1							22
11							2	2	1	4	6	6	9	5	2	4	1	3	1			1			1		48
10[e]						1		1	1	10	7	6	6	1	1	2		3	1		1		1				42
10				1		1	2	3	10	8	4	3	6	3	4	2	4	1									53
9[e]	1				1	1	1	3	3	5	5	11	1	1	2	2		2		1	1		1				42
9				1	2	4	2	6	6	7	6	3	4	3	3	1		1		1							50
8[e]		1	1	1	4	3	3	2	8	6	4	4	1	2	3			1									44
8			2	1	1	3	3	5	3		4	5	2		3		1										33
7[e]		1			2	2	3			1	1																10
7			1		1		3	2	2	4																	13
6[e]		1				1			1																		3
6					1																						1
Total	1	3	5	4	12	20	20	25	41	48	52	54	47	29	34	22	12	22	7	7	3	5	2	0	3	2	480

r=.535

Note—Columns include scores 0–4, 5–9, 10–14, etc. Rows include mental age ranges of one-half year: 6-6 equals 6-6 to 6-12, etc.

Since the group is of limited heterogeneity, a better index of the value of the test can be found by computing the probable error of a mental age based on this one test. Treating separately the mental age distributions of those naming 30 to 39 words, 40 to 49, 50 to 59, etc. we get the following:

TABLE II

Score	Median mental age Stanford-Binet	P. E.
30–39	8 years 10 months	8.25 months
40–49	9 years 0 months	10.7 months
50–59	9 years 11.2 months	9.9 months
60–69	10 years 7.4 months	15.2 months
70–79	11 years 4 months	15.2 months
80–89	11 years 3 months	14.4 months
90–99	11 years 6 months	10.5 months
	Average P. E.	12.01 months

The average P. E. of a mental age derived from the 60 word test is almost exactly one year, or at least this is the case when the number of words named is between 30 and 100. Below 30 and above 100 our subjects are too few to give data of value for this purpose. With a P. E. of 12 months, the test may be expected to give a mental

age correct within 18 months in 69 per cent. of cases, and within 24 months in 82 per cent. of cases. The 60 word test is a somewhat less accurate measure than the vocabulary test, which in the case of children, gave a P. E. of 9½ months in terms of mental age. However, it is not as inferior as would appear from a comparison of the coefficients of correlation with mental age (.91 and .535).

The Reliability of a One-Minute Test

One would naturally expect much greater accuracy for a three-minute than for a one-minute test. To test this assumption the score for the first minute of the test has been separately correlated with mental age. The results are shown in Table III.

TABLE III

Total Words in One Minute

Mental Age	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	Total
17⁶	.	.	..	..	..	..	...	..	..	..	1	..	..	..	..	1
17	.	.	..	..	..	..	...	..	..	..	..	..	..	..	..	0
16⁶	.	.	..	..	..	..	...	..	..	2	..	1	..	..	..	3
16	.	.	..	..	..	1	...	..	..	1	..	..	..	..	..	2
15⁶	.	.	..	..	..	..	1	..	..	..	..	..	..	..	..	1
15	.	.	..	..	..	..	...	2	3	..	..	..	..	..	..	5
14⁶	.	.	..	..	..	..	3	1	1	..	..	..	..	..	..	5
14	.	.	..	..	..	..	1	2	2	..	..	..	..	..	..	5
13⁶	.	.	..	..	..	..	3	6	3	..	..	1	..	..	..	13
13	.	.	..	..	2	2	6	6	4	2	..	..	1	..	..	23
12⁶	.	.	..	..	..	3	3	8	3	3	1	1	..	..	..	22
12	.	.	..	1	3	8	8	10	1	2	1	..	..	..	..	34
11⁶	.	.	1	1	2	2	5	8	..	1	..	..	..	..	..	20
11	.	.	1	3	7	8	9	8	4	3	3	..	..	..	..	46
10⁶	.	.	..	1	2	10	9	5	5	3	..	..	..	..	..	35
10	.	.	2	4	8	14	13	11	3	..	..	..	..	..	..	55
9⁶	.	1	..	3	10	7	11	6	3	1	..	..	..	..	..	42
9	.	.	1	8	5	11	12	7	1	..	..	..	..	..	..	45
8⁶	1	.	5	4	3	9	10	4	1	..	..	..	..	..	..	37
8	.	.	3	4	7	6	7	3	..	..	..	..	..	..	..	30
7⁶	.	.	3	..	6	..	1	..	..	..	..	..	..	..	..	10
7	.	.	1	4	4	3	...	..	..	..	..	..	..	..	..	12
6⁶	.	.	2	1	..	..	...	..	..	..	..	..	..	..	..	3
6	.	.	..	..	..	..	...	..	..	..	..	..	..	..	..	0
5⁶	.	.	..	1	..	..	...	..	..	..	..	..	..	..	..	1
Total	1	1	19	35	59	84	102	87	34	18	6	3	1	..	..	450

r = .52

Note—Columns include scores 0–4, 5–9, 10–14 etc. Rows include mental age ranges of one-half year: 6–6 equals 6–6 to 6–12, etc.

The writer was surprised to find that one minute constitutes as good a measure in this test as three minutes. The correlation with mental age is almost exactly the same in the two cases, (.535 and .52). The P. E. of mental age based on the one-minute test is approximately 13 months; or only a month higher than for the three minute test. The saving of two minutes would be material and it is proposed that the test be shortened to one minute and that 28 words be required for a pass at year X instead of 60 words in three minutes.

The correlation between the first minute and the last minute was computed and found to be .58.

The Influence of Foreign Language in the Home

The correlation with mental age was computed separately for 102 Portuguese and Italian school children. Nearly all of these were from homes in which a foreign language is spoken, although a majority of the children had learned English before entering school. The correlation was .43, which is slightly lower than for American children. The medians for the different mental ages did not, however, differ materially in the two groups. (See Table IV.) The conclusion would be that the test is not seriously enough vitiated by the language factor to justify its omission in testing children of foreign parentage, provided the children have spoken English for three or four years.

Decrease of Words Named in Successive Minutes

Theoretically, the percentage of decrease in number of words named in the successive minutes might be significant. Observation had led the writer to believe that average and superior children would show a smaller percentage of drop from the first minute to the second and from the second to the third, than do dull and feeble-minded subjects. Accordingly the percentage of drop was calculated for 310 children and the results were then correlated with intelligence quotient. This was done for the 180 boys and 130 girls separately.

The results show how erroneous an impression is likely to be when based on nothing but casual observation. The percentage of drop from the first minute correlated with intelligence quotient to the extent of only .227 with boys and .19 with girls. The percentage of drop from the second minute to the third minute correlated with intelligence quotient to the extent of .185 with boys and .135 with girls.

The children were then divided into three groups according to intelligence quotient: (1) those below 80; (2) those between 80 and 110; and (3) those above 110. The median percentage of drop for the three groups was as follows:

	First minute to second minute		Second minute to third minute	
	Boys	Girls	Boys	Girls
IQ below 80	40%	46%	51.2%	54%
IQ 80–110	31.5%	35%	46.7%	41.3%
IQ above 110	33.7%	31.7%	37.5%	43%

Although the coefficient of correlation is low, there is a consistent tendency for bright children to maintain their initial speed considerably better than do the dull. There are so many exceptions, however, that the amount of drop cannot be regarded as a dependable symptom in clinical work.

Incidentally it is interesting to note that the third minute is about as inferior to the second as the second is to the first; also that in percentage of drop there are no significant sex differences.

Increase by Mental Age in the Ability Tested

Table IV shows for each mental age the median number of words named in three minutes by the sexes taken together, by the boys and girls taken separately, and by the Portuguese-Italian group; also the medians for the number named in one minute (American group).

TABLE IV

	Three Minutes			One Minute	
Mental age	Sexes together	Boys	Girls	Port-Italian	American group sexes together
13-6 to 14-5	78.8	77	83.3	81	38.1
12-6 to 13-5	77.5	72.5	79.5	79.	37.1
11-6 to 12-5	69.5	64.1	72	77	33.5
10-6 to 11-5	69.2	68.3	69.4	70.5	32.8
9-6 to 10-5	64.4	60.8	67.5	62.5	30
8-6 to 9-5	56.5	61.1	57.5	56.7	28.5
7-6 to 8-5	49	43.7	45	53.5	23.8
6-6 to 7-5	47.5	47.5	49	50	19

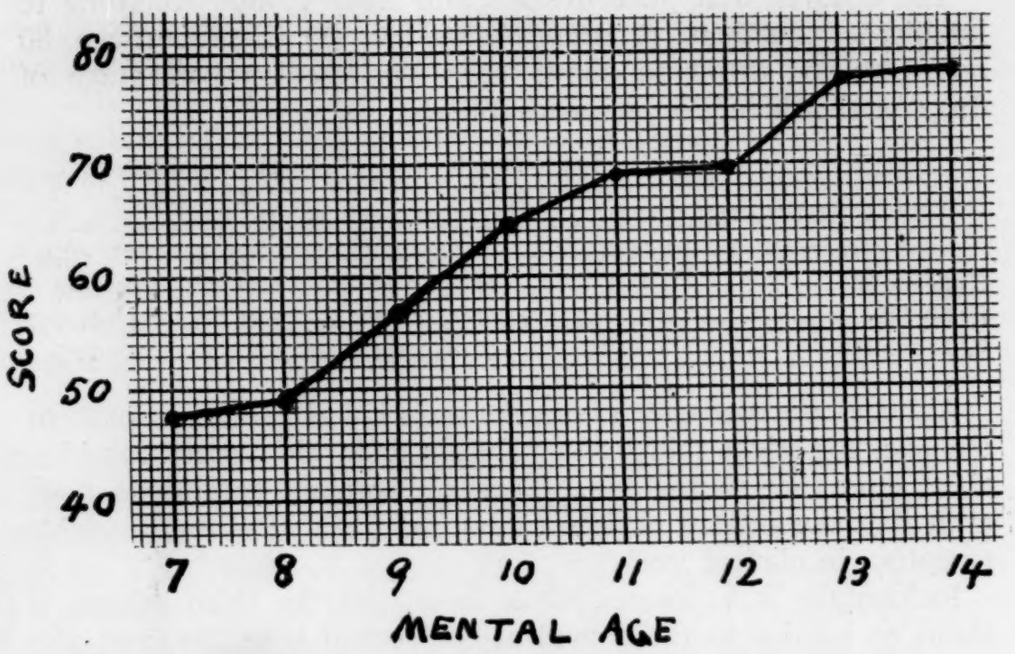

FIGURE I. Mental age increase in number of words named in three minutes. (American children, both sexes.)

DEPENDENCE UPON VOCABULARY

For 360 children the correlation between vocabulary and number of words named in three minutes was .487. The two abilities are evidently fairly closely related. This is probably due to their mutual dependence upon intelligence. It is not known whether children of the same mental age but of widely different vocabulary would differ greatly in the 60 word test.

CRITICISM OF THE TEST

In the first place the writer would urge the desirability of working out a satisfactory method of giving credit for quality as well as for quantity. This is important but will be a difficult task.

Performance in the test is unquestionably influenced by the conditions under which it is given and by the surroundings. In view of the impossibility of providing a standard room, containing always a given equipment, it has been suggested that the subject be required to close the eyes. It would be interesting to see whether this procedure would increase the correlation with mental age.

The performance is also influenced by the amount of emphasis placed by the examiner upon speed. In *The Measurement of Intelligence* the writer has indicated the degree of emphasis which he regards permissible.

Finally, the test is by no means easy to score. When words are named at the rate of 20 to 40 per minute it is hard to keep count, especially since the examiner must also watch the time, note repetitions, and pay some attention to the order of associations and quality of performance. The test is harder work for the examiner, if he is conscientious, than for the subject. As it is ordinarily not possible to record the words unless one can write in shorthand, the record should be made by pencil marks. Even then a class of students recording a demonstration test will often vary in the count as much as three or four words if the subject is rapid. The disagreements are largely caused by the neglect to eliminate repetitions and to confusion in recording compound words. It is recommended that words like "door knob," "sulphuric acid" etc., should be counted as one.

SUMMARY

1. The 60 word test is, for a single test, a fairly good measure of intelligence. In 50 per cent of cases it yields a mental age within a year of that secured by the Stanford-Binet Scale.

2. The time can be reduced to one minute without appreciable loss of accuracy, as a one-minute test yields a mental age whose P. E. is only 13 months.

3. In the case of children who have attended school for three or four years the test is not seriously vitiated by the use of a foreign language in the home.

4. The percentage of drop in the number of words named in successive minutes is greater for dull than for bright subjects, but the correlation is low.

5. The percentage of drop from the first minute to the second is about the same as from the second minute to the third.

6. The ability tested increases fairly regularly in successive mental ages.

7. Special care is necessary to secure accuracy and uniformity of scoring.

A PRELIMINARY TEST IN CHEMISTRY

HANOR A. WEBB

Department of Chemistry, George Peabody College for Teachers

Many types of educational problems are being studied from quantitative data. The value of standard tests in elementary school subjects has encouraged similar experiments in high school classes, and one by one "tests" for the various secondary subjects are published.

A clear distinction has been recently expressed between tests of skill in performance, and tests of range of information.† Performance, both in speed and accuracy, is most conveniently measured by means of a series of problems to be solved within definite time limits—range of information is determined with a somewhat similar series of queries concerning facts which, in general, the student either knows, or does not know. In connection with the subject of chemistry, the distinction between problem and informational tests may be more clearly perceived by comparing the preliminary questions prepared respectively by Professors Jones* and Bell.†

A "standard test" is usually designed to compare the results of previous instruction with some properly derived standard,—measurements more relative than absolute. But little has been done towards the determination of the amount of information possessed by the children *before* they start in upon the period of instruction. These so-called "apperceptive masses," upon which the new concepts must necessarily be built, are undoubtedly of infinitely greater qualitative variety, and wider quantitative range, than even the most scattered results of a term's instruction. The question "What do they already know about the subject?" is in every teacher's mind as he faces a beginning class.

This "tare" in weighing the results of instruction must receive attention if the achievements are to be properly interpreted. But on the qualitative side, what the children already know is the teacher's accessory *before* the fact,—flexibility in method or adaptation of subject matter depends on the discovery of the foundations of information already prepared for the new construction.

---

†J. CARLETON BELL. *A Test in First Year Chemistry*, Journal of Educational Psychology, Vol. 9: 1918, 199–209.

*FRANKLIN T. JONES. *Union Tests in Physics* (Chemistry in preparation), Cleveland, O.

One of the first conceptions necessary in understanding elementary chemistry is the essential difference between *elements, mixtures,* and *compounds.* Their specific properties are illustrated in the first experiments,—their definitions are among the first learned. Texts usually consider these three topics of equal difficulty, and equal time is no doubt given to each in most classrooms. But preconceptions and confusions regarding these substances already exist in the minds of students. Which class of matter is most easy to understand? Which most difficult? Do the experiments of the manual and text clearly illustrate their differences?

For some years the writer has been handing a test sheet on elements, mixtures, and compounds to his normal school and college classes in beginning chemistry. At the head of the sheet, definitions for each class of matter were called for,—then an alphabetical list of fifty well-known substances was given,—nine elements, thirty-two mixtures, and nine compounds required classification. The student filled out the paper without assistance *after* having performed the experiments on this topic in the manual, and studied the text. The sheets were handed in *before* class discussion, it being the idea of the instructor to glance through them, note the prevailing misconceptions, and adjust class demonstrations and discussions accordingly

Out of the accumulation of some years, two hundred fifty papers have been selected, in which the definitions of the three classes of matter are correctly given, and in which all of the fifty substances are classified. The complete data of this test is displayed in Table I

The Results

I. Majority judgments. Table II.

Of 9 elements,	8 correct,	0 incorrect.
Of 32 mixtures,	9 correct,	10 classified as compounds.
Of 9 coumpounds,	8 correct,	0 incorrect.

Majority judgment is correct as to elements. It also judges compounds truly when they are compounds, but is greatly confused by most mixtures, interpreting them as compounds.

II. Plurality judgments. Table II.

Of elements,	1 correct,	0 incorrect
Of mixtures,	3 correct,	10 incorrect.
Of compounds,	1 correct,	0 incorrect.

TABLE I

Rank Order of Substances, as Judged

	Elements		Mixtures		Compounds	
	Iron	201	Lemonade	207	Water	188
	Phosphorous	195	Face Powder	184	Bone	156
	Zinc	190	Chewing Gum	167	Sugar	150
	Sulfur	180	Gunpowder	158	Soda	149
	Diamond	180	Ink	150	Glass	148
	Copper	180	Soil	148	Salt	145
	Lead	172	Soap	129	Hair	144
	Nickel	156	Milk	129	Alcohol	144
	Soot	101	Flour	125	Epsom Salts	143
	Steel	99	Gelatin	122	Lard	142
	Wool	92	Air	116	Coal	141
	Cotton	84	Dyes	113	Eggs	141
Q_3						
	Salt	79	Paper	112	Meat	139
	Granite	77	Butter	110	Rubber	138
	Silk	72	Smoke	100	Wood	138
	Hair	72	Tea	96	Camphor	136
	Brass	69	Camphor	94	Corn	130
	Soda	61	Silk	89	Dyes	130
	Coal	61	Eggs	88	Cotton	123
	Rubber	59	Corn	87	Brass	122
	Chalk	57	Coffee	85	Chalk	120
	Alcohol	56	Granite	82	Coffee	118
	Tea	53	Meat	76	Paper	117
	Wood	51	Lard	74	Soap	117
	Sugar	49	Chalk	73	Steel	115
M						
	Coffee	47	Epsom Salts	72	Butter	115
	Smoke	46	Bone	65	Wool	112
	Glass	43	Wood	61	Gelatin	108
	Meat	35	Glass	59	Air	106
	Epsom Salts	35	Brass	59	Smoke	104
	Lard	34	Soot	57	Milk	102
	Flour	33	Rubber	53	Tea	101
	Corn	33	Sugar	51	Soot	92
	Bone	29	Alcohol	50	Flour	92
	Air	28	Coal	48	Granite	91
	Butter	25	Wool	46	Ink	90
	Paper	21	Cotton	43	Silk	89
	Eggs	21	Water	42	Soil	87
Q_1						
	Water	20	Soda	40	Gunpowder	86
	Gelatin	20	Steel	36	Chewing Gum	78
	Camphor	20	Hair	34	Diamond	65
	Milk	19	Nickel	29	Nickel	65
	Soil	15	Salt	26	Face Powder	63
	Ink	10	Lead	20	Lead	58
	Dyes	7	Zinc	16	Copper	56
	Gunpowder	6	Sulfur	15	Sulfur	55
	Chewing Gum	5	Copper	14	Phosphorus	45
	Soap	4	Phosphorous	10	Zinc	44
	Lemonade	3	Iron	9	Iron	40
	Face Powder	3	Diamond	5	Lemonade	40

Since a substance might be judged as belonging to any of three classes, some did not receive a majority in any class.

The combined majority and plurality judgment places all elements correctly, all compounds correctly, 12 out of 32 mixtures correctly (37.5%), and errs only in classifying 20 mixtures (62.5%) as compounds.

TABLE II

Substances Grouped by Majority and Plurality Judgments

Elements		Mixtures		Compounds	
Iron	201	Lemonade	207	Water	188
Phosphorous	195	Face Powder	184	xBone	156
Zinc	190	Chewing Gum	167	Sugar	150
Copper	180	Gunpowder	158	Soda	149
Diamond	180	Ink	150	xGlass	148
Sulfur	180	Soil	148	Salt	145
Lead	172	Milk	129	xHair	144
Nickel	156	Soap	129	Alcohol	144
		Flour	125	Epsom Salts	143
				xLard	142
Soot	101			xCoal	141
		Gelatin	122	xEggs	141
		Air	116	xMeat	139
		Silk	89	xRubber	138
				xWood	138
				Camphor	136
				xCorn	130
				Dyes	130
				xCotton	123
				xBrass	122
				Chalk	120
				xCoffee	118
				xPaper	117
				xButter	115
				xSteel	115
				xWool	112
				xSmoke	104
				xTea	101
				xGranite	91

The horizontal line shows division between majority and plurality judgments. x indicates incorrect classification.

In discussion, students always question the classification of a few of these substances, where the characteristics are not well defined, or not uniform. For example, soap, though chemically a compound, is commercially a mixture, and is more properly classified as such with beginning students.

TABLE III

Rank Order of Substances as Correctly Judged

M	Lemonade	207	p M	Gelatin	122
E	Iron	201	p C	Chalk	120
E	Phosphorous	195	p M	Air	116
E	Zinc	190	x M	Paper	112
C	Water	188	x M	Butter	110
M	Face Powder	184	p E	Soot	101
E	Copper	180	x M	Smoke	100
E	Diamond	180	x M	Tea	96
E	Sulfur	180	p M	Silk	89
E	Lead	172	x M	Eggs	88
M	Chewing Gum	167	x M	Corn	87
M	Gunpowder	158	x M	Coffee	85
E	Nickel	156	x M	Granite	82
M	Ink	150	x M	Meat	76
C	Sugar	150	x M	Lard	74
C	Soda	149	x M	Bone	65
M	Soil	148	x M	Wood	61
C	Salt	145	x M	Glass	59
C	Alcohol	144	x M	Brass	59
C	Epsom Salts	143	x M	Rubber	53
C	Camphor	136	x M	Coal	48
C	Dyes	130	x M	Wool	46
M	Milk	129	x M	Cotton	43
M	Soap	129	x M	Steel	36
M	Flour	125	x M	Hair	34

The letters E, M, and C, stand for the correct classification as element, mixture, or compound.

All substances in the first column are correct by majority judgment.

In second column, p indicates correct by plurality judgment; x indicates incorrect by plurality judgment.

III. Correct judgments. Table III.

This table represents the "order of familiarity" with the composition of these fifty substances. It is again to be noted that the elements in the list rank high in familiarity. If unfamiliar elements had been chosen, their rank would have been low, but such a selection would have violated the spirit of this test,—all of the fifty substances are familiar objects *per se*, and the variation is purely in the knowledge of their general composition.

The majority of mixtures received few correct judgments.

IV. All judgments. Table I.

Table I. gives three ranked distributions which admit of much statistical treatment. Certain values are perhaps significant in this study of the misconceptions in the minds of students regarding the three classes of matter.

(*a*) *The median.*

Median number of judgments for each classification: elements, 48; mixtures, 72.5; compounds, 115.

Students were more decided in their judgments concerning elements, for half of the substances received 48 or fewer votes. The judgments are most scattered when compounds are considered.

(*b*) *The average.*

Average number of judgments for each substance classified: as element, 64.1; as mixture, 77.5; as compound, 108.3.

The average is a better central tendency to consider in this problem than the median, for the extreme cases are here significant. The students have shown a decided tendency to classify substances as compounds.

(*c*) *The quartile ranges.*

Range of	for 9 elements	for 32 mixtures,	for 9 compounds.
Q_4	1830	1784	1791
Q_3	816	1166	1643
Q_2	430	706	1289
Q_1	132	254	695
Total judgments,	3208	3874	5418
Correct judgments possible.	2250	8000	2250

The significant values here are not so much the quarter points, as the number of measures (judgments) in each quartile range. The concentration of judgments in the upper ranges of the element group is in marked contrast to the scattering throughout all ranges of the compound group. The element column denotes a confidence and definiteness of opinion; the compound column indicates indecision,—the feeling, "If I don't know what it is, I'll call it a compound." And from this, the mixtures have suffered most. Not only have plurality judgments shown the tendency to classify mixtures as compounds, but this tendency is here shown to be of

major proportions. Of the 8000 correct judgments which the mixtures might have received, 3618 went *unpaired* into the compound column.

(*d*) *The extreme measures.*

No. of judgments,	elements,	mixtures,	compounds.
Highest,	201	207	188
Lowest,	3	5	40

The principal confusion is again shown to lie in the compound class. That only 188 (75.2%) correctly understood the nature of water, and that 40 (16%) erroneously classified iron and lemonade as compounds, shows that the prevalent conception is most hazy.

V. Measures of relationship.

(*a*) *Nature and per cent. of errors.* Data from Table I. shows the following relationships of true and erroneous judgments.

Substance	Classified as		
	element,	mixture,	compound
Elements,	69.1%	7.8%	23.1%
Mixtures,	15.9%	39.2%	44.9%
Compounds,	17.1%	24.9%	58.0%

Thus, elements are the best, and mixtures the least understood. The largest error is in classifying mixtures as compounds,—the smallest error, classifying elements as mixtures. The intermediate errors are of moderate size.

(*b*) *Coefficients of correlation.*

As the judgments of two hundred fifty students has given each of fifty substances a rank in three separate classifications, the correspondence of these ranks, (taking into full account the actua number of judgments for each substance), is best expressed by the coefficient of correlation. The best correlation in these cases would be a decidedly minus value, for if many substances were correctly judged, they should rank high in their own classification column, and low under each of the other heads. A strongly negative correlation shows that the properties which distinguish two classes are clearly understood by students, while a smaller negative numerical value shows that there has been much haphazard swapping of judgments. The coefficients of correlation*

between elements and mixtures is −.790
between elements and compounds, −.555
between mixtures and compounds, −.063

*Calculated by "Pearson's formula." The rather complex method is clearly outlined in *Statistical Methods Applied to Education*, by H. O. RUGG, pp. 251–270.

CONCLUSIONS

In learning to distinguish between elements, mixtures, and compounds, there is one outstanding, predominent error of judgment made by beginners in chemistry,—the classification of mixtures as compounds. An instructor must earnestly supplement the experiments of manual and text at this point. Other misjudgments are no greater in quantity than would be expected of a normal class.

At least two important causes of error are apparent:

(a) The student has relied on the physical appearance of *homogeneity*,—glass, brass, rubber, steel, etc., seem to be uniform in structure, like compounds. The typical experiments, with iron filings and sulfur, sugar and sand, etc., have been inadequate to clear up this point.

(b) Most students have little conception of the cell structure of plant and animal tissues. Meat, bone, wood, hair, etc., being cellular, are necessarily mixtures.

This study emphasizes the great need for a more scientific attitude in relation to the common things of daily life, and is possibly an argument for some type of science instruction in the grades. That less than half of two hundred fifty nearly grown young men and women knew that air was a mixture makes us wonder how simple a scientific fact must be to become a matter of common knowledge under present methods of instruction. Teaching the three classes of matter is but one of the many problems of beginning chemistry where it is well to know the extent of previous knowledge, right and wrong, and if the standard test, arranged for a given subject, can portray to us the content of a student mind "after taking," let efforts be made to complete the contrast, and show us what was there "before."

COMMUNICATIONS AND DISCUSSIONS

THE EFFECT OF THE SUMMER VACATION ON ABILITY IN THE FUNDAMENTALS OF ARITHMETIC

This investigation was undertaken primarily for the purpose of ascertaining the influence that the summer vacation has on arimetical ability, with a view to locating weak points needing emphasis on return to school. Inasmuch as the school has a cosmopolitan population including children of varying economic and social environment, three distinct types of vacations were recognized; the vacation spent in play, the vacation spent in work and the vacation spent in study. The comparative influences of these three types of vacations on the ability in fundamentals were studied.

The measuring instrument used was the Courtis Test in Fundamentals, Series B. Form 1 was used in June 1918 and Form 2 was used in September 1918. All four tests, that is, addition, subtraction, multiplication and division were given.

Terms Used

"Play Group"—Children whose vacation was passed doing whatever the child desired, but not in working or studying to make up a grade.

"Work Group"—Those children whose vacation was passed in employment. Unless the child worked at least four weeks the vacation was not classed as "work." A child employed to help parent in store, etc., for only two or three hours a day was not considered in this group.

"Study Group"—Those children whose vacation was spent in studying at a summer continuation school to make up a grade.

Scope of Test

The test was given to all 5A, 5B, 6A, 6B, 7A, 7B children in Public School 9, The Bronx. 861 children of whom about 1–3 were girls were tested in June. Owing to transfers the number of these same children in September was reduced to 747. The "work" and "study" groups remained constant, only those children being included who took the test both in June and September. In tabulating the results no distinction was made between boys and girls, because no attempt was made to study the relative ability of the sexes.

Conduct of the Test

The children were taken into an assembly room seating eight classes at a time. The tests were all conducted by the writer personally. The directions that appear at the top of each test sheet were read according to instructions and the prescribed time given for work. All four tests were given at the same session, intervals

of a minute or two being allowed between tests for stretching and relaxation. The entire test consumed about forty minutes for each group of classes.

The weather on all occasions in June and September was about the same; fairly warm but comfortable. In September the humidity was somewhat greater than in June making it slightly oppressive to some of the children. The difference, however, was not sufficient to make it a consideration.

The tests were corrected by a 6B class trained for the purpose and under the supervision of the writer and the class teacher.

Each class teacher tabulated her class results according to instructions given by the writer. A conference was held for the purpose. Those tabulations were checked by the writer and then summarized by him.

Although the primary interest was in the comparative abilities in June and September of the "work," play" and "study" groups the results were also summarized by grade groups so that a grade standard could be derived for purposes of comparison. In the present study the grade groups were combined into year groups as that was deemed more satisfactory for the purpose in view.

Unfortunately the "work" and "study" groups in each year did not prove large enough to warrant comparisons being made, so that it was necessary to tabulate the "work" and "study" children of all grades into one total, giving one "work" group of 134 children and one "study" group of 37 children for the three years combined. The year groups of course include the "work" and "study" children as well as the "play" children but the two former are comparatively so few in number that in any one grade they affect the results very little, so that for purposes of comparison the year group as given may be considered as a "play" group.

Results

Although we are here concerned with the influence of the vacation on the results in the various groups only, it is of interest to note the distribution in the different tests. From a study of the tables of distribution which were derived but are not given here for want of space, several things are clear. In speed there is a wide distribution in all groups, being particularly marked in the seventh year and being of wider range for the addition and subtraction tests than for the multiplication and division tests. So marked is this range of distribution that at times we find one third of the fifth year group showing a better rate than one half of the seventh year group and, conversely, about one third of the seventh year group showing a rate poorer than one half of the fifth year group. This wide range in rate exhibited by all the groups would seem to indicate the necessity for special drill for those pupils in the lower end of the scale, and for relief from drill of those children at the upper end of the scale. That this is not done more often than it is done at present is due to difficulties inherent in the crowded condition of our classes.

TABLE OF MEDIANS AND VARIATIONS FROM MEDIANS

RATE

GROUP	TEST I			TEST II			TEST III			TEST IV		
	June Median	Sept. Median	Median Variation	June Median	Sept. Median	Median Variation	June Median	Sept. Median	Median Variation	June Median	Sept. Median	Median Variation
7th Year	11.1	12.7	+1.6	13.1	12.9	− .2	10.5	10.5	.0	10.1	9.9	− .2
6th Year	8.7	10.6	+1.9	11.5	11.0	− .5	9.7	8.8	− .9	8.9	8.8	− .1
5th Year	8.6	8.8	+ .2	10.4	9.6	− .8	8.2	7.9	− .3	6.5	6.4	− .1
Total "Play" Group	....	....	+3.7	....	....	−1.5	....	....	−1.2	....	...	− .4
Average "Play" Group	....	....	+1.2	....	....	− .5	....	....	− .4	....	...	− .1
"Work" Group	10 7	11.1	+ .4	12.2	11.7	− .5	10	9.7	− .3	9.6	9.0	− .6
"Study" Group	9.7	11.2	+1.5	12.0	11.3	− .7	9.4	8.8	− .6	9.0	7.9	−1.1
ACCURACY												
7th Year	81	69	−12	93	82	−11	85	73	−12	89	84	− 5
6th Year	78	71	− 7	88	87	− 1	83	74	− 9	89	83	− 6
5th Year	75	70	− 5	87	88	+ 1	81	70	−11	79	70	− 9
Total "Play" Group	..	..	−24	..	..	−11	..	..	−32	..	..	−20
Average "Play" Group	..	..	− 8	..	..	− 4	..	..	−11	..	..	− 7
"Work" Group	79	70	− 9	88	86	− 2	81	71	−10	81	81	0
"Study" Group	80	70	−10	91	86	− 5	81	72	− 9	88	84	−4

Except in the matter of distribution, the accompanying table gives a summary of all the results in terms of medians and variations of medians from June to September. A study of these enables us to summarize the results as follows:

1. In speed there is a general decrease from June to September, except in addition. This decrease, however, never amounts to more than a year's work. The "work" group as a rule presents a more favorable showing than the "study" group and as good or better than the "play" group except in multiplication.

2. In accuracy there is a decided decrease in efficiency in all groups. This decrease averages the equivalent of almost two years' work with slight exceptions in Test 2 and Test 4. The highest median in September is lower than the lowest median in June. The worst showing is made by the seventh year group which shows the highest median in all tests in June and the lowest or next to the lowest in September (except in division). The loss of efficiency in this group at times amounts to more than the equivalent of two years' work, the September median in addition, for example, being lower than the June median for the same test of the fifth year. The "work" group as a rule shows more favorable results than the "study" group.

3. When taken in conjunction with the increased rate, the decrease in accuracy in addition is significant. "Familiarity has bred contempt." Although the children have probably been drilled more often in addition than in the other fundamentals and thus have acquired a facility in handling such examples, the difficulties of the process have not been given close attention. The many processes involved, ("carrying," the "attention span," and other difficulties) have not been given the conscious attention necessary for the formation of a habit. In the fifth year, where the habit is of recent formation, the loss in accuracy is not so great and the gain in speed is negligible.

4. By comparison with the "play" group (averaged for the 5th 6th and 7th years) the "study" group shows the greatest loss in rate (except in addition, where it shows the greatest gain) while the "work" group is as good or better than the "play" group, except in division. In accuracy the "work" group shows a much smaller loss than the "play" group while the "study" group shows a greater loss in addition and subtraction and a smaller loss in multiplication and division.

Conclusions

Though the vacation brings about a general loss in efficiency both in speed and in accuracy, the loss in speed is such that it may be neglected in an attempt to make up the loss in accuracy. On return to school the emphasis, for a time at least, should be on accuracy to the exclusion of speed. In addition, particularly, accuracy should be emphasized. In all fundamentals those pupils

showing proficiency should be excused from drill whereas those showing particularly poor results should be placed in special drill classes.

The comparatively better showing of the "work" group and the poor showing of the "study" group would seem to indicate that the child who keeps his brain alert by meeting actual problems of life is in a better position to retain what he has learned than the one who plays or studies. The poor showing of the "study" group, however, may be partly explained by the fact that the studying group did not continue to the opening of school. The last four weeks of the vacation were spent in play by this group.

The great loss in efficiency of the "play" group would seem to point to the advisability of an all year school for the child. Whether this conclusion is justified, however, would depend upon the power for recovery shown by each group. This would have to be the subject of further tests.

MAURICE A. GARFINKEL.

Public School 9, The Bronx,
New York City.

A COMPARISON OF THE RESULTS OF GROUP TESTS AND THE POINT SCALE TEST

Every year mental measurements assume a greater importance in this country. As public school teachers come to have a scientific knowledge of applied psychology so that they can use psychological methods and understand and interpret the results of mental measurements, such measurements will come into general use. Hitherto the various adaptations of the Binet Scale have been used almost exclusively, in the practical diagnosis of subnormality. But the use of the Binet tests or any other individual tests for all the pupils of a school system is now, and perhaps always will be, an impossibility. However, careful annual testing of *all* the children in our schools is highly desirable. There is, therefore, a present demand for the extension, standardization, and perfection of group tests. If group tests give practically the same results, measure the same important functions, then there is no reason why they should not be used instead of the individual tests except in the case of the extreme variants, who should, of course, have a much more extensive examination than is possible with a few group tests.

The writers have undertaken a comparison of the results obtained by use of the Yerkes-Bridges Point Scale with the results obtained by the use of standardized group tests. The subjects were eighty public school children in grades four, five, and six. The Yerkes-Bridges tests were carefully given to the children dur-

ing the fall and winter of 1917-18. We began with the higher grade and proceeded to the lower. The group tests used were the seven mental tests standardized by one of the writers (W. H. P.).* They were administered simultaneously to all the children of each grade. The individual tests were all given, and the results of all the group tests were graded by one of the writers (G. E. B.). Absolute uniformity was therefore secured in the grading and evaluation of the results. The correlation by the Pearson formula between group test rating and Point Scale rating was .619; P. E. .046. If instead of using the absolute marks of the pupils, we use their ranks and compute the correlation by the Spearman foot rule formula, we get R = .84.

The pupils of each of the three grades were then ranked on the basis of the teachers' judgments of their ability. The correlation between group test ranking and the teachers' ranking, obtained by the use of the Spearman Foot-Rule formula, were: fourth grade, .565; fifth grade, .71; sixth grade, .80. Correlations by use of the same formula between Point Scale ranking and teachers' ranking were: fourth grade, .42; fifth grade, .95; sixth grade, .655. The average correlation, therefore, between the group test ranking and teachers' ranking was .692; and the average correlation between Point Scale ranking and teachers' ranking was .675. These correlations are practically the same, that from group tests being .017 the higher. Therefore, whatever value inheres in the Binet Scale as a means of measuring intelligence, the same value inheres in the group tests. By every criterion that it was possible to apply, group test rating was found to be as accurate as Binet test rating.

If the raw correlation by the Pearson formula is .619 and by the Spearman formula .84, the true correlation between the two sets of measurements is probably near unity. Nor should it be surprising that such is the case, for the Binet tests, as well as the group tests, are merely measures of the various aspects of learning capacity, the efficiency of the associative processes, and the effectiveness of memory, that is to say, in both cases we are probably measuring the same mental functions. The only question is as to the most accurate method of measuring these functions. Our results seem to show that the group test method is as accurate as the individual test method.

The significance of this conclusion is, we think, great. When we further perfect and standardize our group tests, the teachers of our public schools will have at their command a fairly accurate method of measuring the mental development of their pupils. By combining the results of a series of group tests—tests that can be given in about an hour—the teacher can, within a few days after

*_The Examination of School Children_, The Macmillan Co., 1913. _Manual for the Mental and Physical Examination of School Children_, University of Missouri Extension Bulletin, No. 21, 1916. _The Science of Human Nature_, Silver, Burdett & Co., 1917.

the beginning of her term of school, have a fairly accurate measure of the native capacity of the children in her grade, and can rank them with reference to their ability as accurately as she can on the basis of a year's experience with them. Of course, we realize that a single measure of rote memory, logical memory, or learning capacity is no more—nor less—accurate than a single measure of jumping capacity, running capacity, or any other physical ability. However, if single measures of the various important mental functions are made and the results pooled, since each is to some extent a measure of general mental development or intelligence, their combined results have considerable validity. Moreover, the tests of the several mental functions can be repeated till, let us say, four of each have been given, at the expense of much less time and energy than could possibly be the case in the use of individual tests. By such repetitions we would not only have a fairly accurate measure of general intelligence by combining the results, but we would have a tolerably accurate measure of each of the several functions.

The accurate gradation of the pupils is of the highest importance not only to the school but to the tax payers who support the schools. In a recent survey of the mental development of the pupils of an entire school system (Webster Groves, Mo.) the children were found to be scattered through the grades with little reference to ability. The frequency surface of fourth grade ability was found actually to overlap to some extent the frequency surface of high school ability. From every point of view this great variation of ability in the same grade is highly undesirable. The school children should be put into small groups of fairly even ability, and each group allowed to proceed in its mastery of the work outlined in the course of study as fast as the group's ability warrants. It is not economical, even from the point of view of the tax payer, to have a pupil repeat a grade. Accurate gradation and proper methods of teaching will make such a thing as a repeater an impossibility. Pupils should not repeat grades, but pass through grades at various rates of progress depending upon ability.

Mental tests in the future will be found helpful in the work of accurate gradation on the basis of ability. This conclusion is no longer a mere theory or opinion, for numerous studies in this country and in England have shown beyond doubt the validity of mental measurements. The present study does, we hope, throw some light on the relative validity of group tests.

G. E. BREECE
W. H. PYLE

University of Missouri.

ABSTRACTS AND REVIEWS

E. A. DOLL. *Clinical Studies in Feeble-Mindedness.* Boston: Richard G. Badger, 1917. Pp. 232. $2.50.

For ten years Dr. H. H. Goddard with a corps of able assistants has been studying the problems of feeble-mindedness in a most careful and scientific way at the Vineland Research Laboratory. When he began his researches relatively little thought was given this subject; now, feeble-mindedness is considered a profound sociological and economical problem and, as a consequence, almost every body from women's clubs to "Uncle Sam" is talking about it. Such confidence, of late, has been placed in the students of feeble-mindedness that the educationalists are demanding of them how to classify children in school; judges to tell them what to do with offenders; eugenists what people should marry; and "Uncle Sam" has commandeered hundreds of these students to tell him whether his boys are fit to fight for freedom.

The Vineland Laboratory is responsible for initiating this great movement and it has done more real research work, collected and evaluated more data, and, in general, contributed more toward giving to students of feeble-mindedness their standing today than any other institution in the world. Therefore it is very fitting that some one representing this laboratory should tell the outside world what conclusions, in reference to feeble-mindedness, they have come to, and something of the procedure followed by them in coming to these conclusions.

This has been done, unusually well, by Mr. Edgar A. Doll in his book entitled *Clinical Studies in Feeble-Mindedness.* Mr. Doll was, for five years, assistant psychologist in the Vineland Laboratory and is now serving as a lieutenant in the department of psychology in the army.

In this book, Mr. Doll has given to the public the important findings which a trained mind can glean relative to feeble-mindedness by living with the feeble-minded and by studying them under the most favorable circumstances. The book is wonderfully condensed—intended primarily for the student of feeble-mindedness—yet it is so readable that it cannot fail to be of great interest and value to the layman. Educators, doctors, lawyers, in fact, anyone who wishes to know the present status of our knowledge relative to feeble-mindedness will find this book not only a rich store house of information, but the many explicit references given by the author make it easy for anyone to pursue the subject further if desired.

The book is divided into two parts. Part I deals with diagnostic criteria and is divided into three chapters. The first chapter sets forth in a clear cut way the problems to be considered. The scope of chapters two and three can best be judged from the following table of contents:

Chapter II, Criteria and Definitions. The Psychological Criterion. The Social Criterion. The Pedagogical Criterion. The Medical Criterion. The Somatic Criterion. The Hereditary Criterion.

Chapter III, Diagnostic Methods and Values. General Principles. The Diagnostic Syllabus; Personal History; Clinical Examination. The Family History Examination; Method; Value for Diagnosis; Value for Etiology; Types of Mating and Expected Offspring. The Social Status Examination; Method; Value. The Medical Examination; Method; Value. The Somatic Examination; Method. The Pedagogical Examination; Method; Value. The Psychological Examination; Importance; Methods; Value.

Part II is taken up with the consideration of six "Illustrative Cases."

The tests used and studies made in family history and other sources of information are recorded in full, and are accompanied by a diagnosis and a prognosis. The treatment of these cases is most illuminating. Contrary to the commonly accepted opinion that the Binet tests are about the only criterion used in passing upon a person's mentality, Mr. Doll in testing these six cases used about twenty-five other tests. He gives the reaction of each case to each test and designates its value in the final diagnosis. The following are the cases considered:

Case I. Obviously Feeble-Minded by Binet-Simon Tests.
Case II. Potentially Feeble-Minded.
Case III. Potentially Normal.
Case IV. Deaf-mute, Neurotic or Psychopathic but not Feeble-minded.
Case V. Psychopathic with slight Intellectual Retardation, not amounting to Feeble-Mindedness.
Case VI. High Grade Borderline Immigrant.

These cases are not intended to be typical but the method of arriving at a diagnosis in each case is typical of the method of procedure that should be followed before making a final diagnosis. The book closes with a carefully selected bibliography of seventeen pages followed by a glossary of the more common technical terms used in the text, and a brief description of thirty tests often used in making mental examinations.

J. M. McCallie.

Supervisor Special Classes,
Trenton, N. J.

A. J. ROSANOFF, HELEN MARTIN, AND ISABEL R. ROSANOFF. *A Higher Scale of Mental Measurement and Its Application to Cases of Insanity.* Psychological Monographs, Vol. 25, No. 3, Whole No. 109. Princeton, N. J.: Psychological Review Company, 1918. Pp. 113. $1.50.

The authors have devised a test on the order of the well-known Kent-Rosanoff association test, but in which the list of 100 stimulus words is drawn from technical terms such as would be encountered in high school and college studies. They have then obtained the simple verbal reactions to these words of 1000 persons of at least a full collegiate education and of 100 noted men of science (many of them in the starred list in American Men of Science). In use the value of each of the 100 responses of a given subject is obtained by reference to frequency tables based on these 1000 or 100 persons and the sum of these values is compared with the average sum for the persons of collegiate training or preferably with the average sum for the noted men (which is referred to as the "high standard" value.) Norms are given for persons of elementary school or high school education and a report is made of the responses of 205 neuropathic individuals. The outcome of the test is influenced not only by education, but also by physiological maturation, sex and temporary condition of the examinee. Without seeking to enter upon a critique of this investigation, the question may be raised whether the method here proposed is not a clumsy, lengthy and confessedly unreliable method of arriving at information that can be gotten by various tests already in the market much more precisely and in a fraction of the time.

G. M. WHIPPLE.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except July and August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Board of Editors
WILLIAM CHANDLER BAGLEY, CARL E SEASHORE,
J. CARLETON BELL, Managing Editor, GUY MONTROSE WHIPPLE.

COLLABORATORS

Sydney Alrutz,	H. H. Goddard,	Will S. Monroe,	L. William Stern
J. R. Angell,	Joseph Jastrow	E. C. Moore,	G. M. Stratton,
J. W. Baird,	G. E. Johnson,	M. V. O'Shea,	G. D. Strayer,
Bird T. Baldwin,	E. A. Kirkpatrick,	W. B. Pillsbury,	E. L. Thorndike,
W. Van Dyke Bingham,	E. H. Lindley,	S. H. Rowe,	E. B. Tichener,
E. F. Buchner,	Anna J. McKeag,	Sante de Sanctis,	J. B. Watson,
W. H. Burnham,	D. P. MacMillan,	E. G. Sanford,	W. H. Winch,
Ed. Claparede,	Adolf Meyer,	W. S. Small,	H. K. Wolfe,
S. S. Colvin,	J. B. Miner,	E. D. Starbuck,	A. H. Yoder.
W. F. Dearborn,			

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated.

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of ten monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X JANUARY, 1919 NUMBER 1

EDITORIAL

MENTAL MEASUREMENTS IN EDUCATION

President Wilson has said that the war has brought forth some very terrible things, but it has also given rise to some very beneficent and valuable things. He was speaking particularly of the conditions in Europe, but the statement applies to this country as well. One of the incidental by-products of our military activity has been the stimulation of popular interest in mental measurements. When the war broke out the psychologists immediately mobilized and offered their services to the government. Not only that but they devised and induced the government to adopt a plan for the testing of recruits that was unlike anything that the world had ever seen before. It is too soon to say just how much this psychological testing and the utilization of the results contributed to increase our military efficiency. Perhaps we shall never be able to tell. But the mere fact that recruits were given mental tests, that in such a life and death matter as waging war the laboratory psychologist had something to offer that was of practical significance, caught the popular fancy and stirred the im-

agination of the man in the street. Never before has the recondite subject of psychology been held in such high esteem. Never before has the daily press devoted so much space to the possibilities of mental measurements. Never before has there been such a favorable opportunity to extend the use of mental measurements in education.

In a circular letter to city superintendents Commissioner Claxton adverts to the fact that many cities have already established bureaus of educational research, and asserts that every city of more than 30,000 inhabitants should have such a bureau. He calls the attention of school authorities to the two or three hundred young men who have been working in the psychological division of the army, and are now about to be discharged, and stresses the "unusual opportunity for city schools to obtain the services of competent men as directors of departments of psychology and efficiency, for such purposes as measuring the results of teaching and establishing standards to be attained in the several school studies, applying mental tests and discovering mental aptitudes of pupils, discovering defective children and children of superior intelligence, and investigating various other vital questions necessary to establish an intelligent basis for promotions, class organization, and special schools."

The suggestion is an excellent one, and it is to be hoped that many school systems will heed the call, and will establish bureaus or enlarge those already established. The time is indeed opportune to press the matter of educational research in each community, and the popularization of the role of psychology in the war tends to create a favorable attitude to such work on the part of the public. But even before the war there was no dearth of trained men and women in our schools capable of directing such bureaus and carrying on such research. The trouble has been that superintendents, boards of education, and the public at large have not been sufficiently convinced of the practical value of mental measurements in education to support them financially and to assign competent teachers to conduct them. It is now incumbent upon all those who believe in the value of scientific studies in education to engage in a vigorous and aggressive campaign to arouse public opinion to demand such studies in the schools as a matter of enlightened and progressive educational policy.

J. C. B.

NOTES AND NEWS

At the December meeting of the New York Society for the Experimental Study of Education the general topic for consideration was "Problems in the Teaching of Civics." Associate Superintendent John L. Tildsley spoke on "Some Needs in Civics Teaching," stressing the need for better teachers, for more time devoted to the subject, and for a more vital connection with the life activities of young people. Professor David Snedden, of Teachers College, in "Practical Problems in Civics" presented two problems for experimental investigation: First, in what respects are selected groups of citizens now failing in citizenship, and what means of education can be devised to supply these defects; second, to what extent can and should the civic virtues of the school community be carried over to the civic community? Miss Mabel Skinner, of the Washington Irving High School, taking as her topic "Community Civics: What Next?" contended that civics should be treated as a laboratory subject, and described the experiments she had been conducting. Mr. Eugene B. Gartlan, principal of Public School 42, The Bronx, discussed "Civics Teaching in Evening Schools," and narrated the difficulties of teaching civics to Russian Bolsheviks in New York City.

At the University of Cincinnati Professor Ernest L. Talbert is giving a course of lectures on *Race Psychology*. Professor Guy A. Tawney, one on *The Nations at War* and Professor Harris M. Benedict, one on *Evolution—Its Nature and Effect upon Human Ideals*.

At the Baltimore meeting of the American Psychological Association Major Bird T. Baldwin presented a paper on *The Function of Psychology in the Rehabilitation of the Disabled Soldier* in which he discussed the aim and scope of psychological service and reported detailed studies in the psychology of voluntary movement. The paper was illustrated with lantern slides.

Professor Samuel C. Kohs, of Reed College, Portland, Oregon, has devised an ingenious instrument for determining the Intelligence Quotient in cases of mental measurement. He calls it the Reed Intelligence Quotient Slide Rule. It consists of a movable cardboard disc with a chronological age scale on its edge centered over a circle with a mental age on its inside and an intelligence quotient scale on its outside. A celluloid indicator is set at the appropriate mental age, the central disc is rotated until the chronological age falls under the hair line, the indicator is then moved to the "I. Q. Line" and the Intelligence Quotient is read directly from the outside scale. With practice any I. Q. can be determined within 15 seconds, and 100 of them read in 25 minutes. The slide rule may be procured from the Reed College Cooperative Store, postage paid, for thirty-five cents.

District Superintendent Arthur C. Perry, Jr., of New York City, recently gave twenty questions on the War to one thousand pupils of the eighth-grade graduating classes. The results were as follows:

1. What do the letters "W. S. S." stand for? 99.5%
2. Who is the commander of the American forces in France? 94.2%
3. Name an illegal use Germany made of her submarines 89.9%

4.	In what month and year did the U. S. enter the war?	89.2%
5.	Name a battle in which U. S. troops took part	84.4%
6.	Upon what nation did Austria first declare war?	80.6%
7.	What nation was the first to declare war?	78.3%
8.	What do the letters "A. E. F." stand for?	73.5%
9.	Name three European nations which were not in the war	70.4%
10.	With what nation did Germany make peace in February, 1918?	70.1%
11.	In what battle in the first year of the war were the Germans stopped in their advance on Paris?	66.4%
12.	With what nation did Germany intrigue, promising to help it invade the United States?	63.4%
13.	By what nation were the German colonies conquered?	61.7%
14.	In what continent were the chief German colonies?	59.4%
15.	About how many soldiers did the United States send to Europe?	58.5%
16.	Name a legal use a nation may make of a submarine in war	58.2%
17.	In what month and in what year did Austria declare war?	40.9%
18.	In what great naval battle were the Germans repulsed?	38.6%
19.	Italy had an agreement to fight with Germany and Austria. Why did she fight against them?	36.7%
20.	With what nations is the United States at War?	31.2%

Among the sub-committees appointed by the Commission on Secondary Education of the North Central Association is one on Educational and Vocational Guidance, the chairman of which is Professor Guy M. Whipple, of the Carnegie Institute of Technology.

During the illness of Professor Herman Harrell Horne, of New York University, three of his classes are being conducted by Dr. J. Carleton Bell, of the Brooklyn Training School for Teachers.

At the recent scientific meetings in Baltimore Colonel Walter D. Scott was elected president of the American Psychological Association, and Major Robert M. Yerkes vice-president and chairman of Section H (Anthropology and Psychology), of the American Association for the Advancement of Science.

Professor William Harry Heck, of the University of Virginia, died of pneumonia following influenza, January 4, 1919, at the age of 39 years. Professor Heck contributed extensively to the experimental study of education, and his book on *Mental Discipline and Educational Values* is an excellent summary of the arguments on this question. He published many papers on *Mental Fatigue in Relation to School Work*, and at the time of his death was preparing an *International Source-Book on Home Study*.

Mrs. Maria Kraus-Boelte, well known for her manifold activities in the spread of the kindergarten movement, has died in New York City at the age of eighty-two years.

Dr. C. Ward Crampton, director of physical training in the schools of New York City, has been granted a year's leave of absence with half salary by the Board of Education for the restoration of his health.

Dr. Raymond Dodge, professor of psychology in Wesleyan University, has been made lieutenant commander in the United States Navy, in recognition of the special psychological tests devised by him for the use of the Navy during the war.

PUBLICATIONS RECEIVED

ROSE LUCIA. *Peter and Polly in Autumn.* Cincinnati: The American Book Company, 1918. Pp. 176. Forty-eight cents.

A delightful story that makes a strong appeal to the little folks. Excellent supplementary reading for the second and third grades.

HELEN MAROT. *Creative Impulse in Industry.* New York: E. P. Dutton and Company, 1918. Pp. xxii, 146. $1.50.

This is a book which challenges the thoughtful attention of every one interested in industrial education. It is characterized by ideas, by grasp of fundamental principles, by earnestness and enthusiasm, by a vision of revolutionary possibilities in the relation of education to industry. In the period of history upon which we are now entering industrial questions are certain to be of the profoundest importance for social welfare. Is the statement made by employers true that from 95 to 99 per cent. of the working force is without productive impulse? If it is now true, is it due to the limitations of human nature and the unalterable demands of modern industry, or is it due to faulty organizations of industrial enterprises? If the former is true, then there is little hope for an industrial democracy. But the author holds that the latter is the real explanation, and looks to education to provide the stimulus for the creative impulse in industry. In the final chapter an extremely interesting experiment in industrial education is outlined, and we are given to understand that there is a probability that the project will be given a fair trial. If so, it will be one of the most valuable and significant contributions that has yet been made to the subject of industrial education.

EUGENE A. NIFENECKER. *Measurements in Spelling* New York: Department of Education, Division of Reference and Research, Publication No. 16, 1918. Pp. 88.

This valuable study of spelling was undertaken at the request of District Superintendent James J Reynolds, in charge of the schools on Staten Island. Twenty-five words from Group T of the Ayres Scale were spelled by each of 5260 pupils in Grades 5A to 8B inclusive. Each word thus received almost four times as many spellings as in Ayres' original study. All grades except the fifth attained higher percentages than indicated on the Scale. The results are tabulated by schools and grades, the age-grade relations are plotted, sex differences are brought out, the home language factor considered, and the factors of time allotment, size of class, and departmental teaching are carefully weighed. A valuable phase of the study is the appendix giving the proportion of correct spellings on each word for each sex of each half grade separately.

NAOMI NORSWORTHY AND MARY THEODORA WHITLEY. *The Psychology of Childhood.* New York: The Macmillan Company, 1918. Pp. xix, 375. $1 60.

This book is intended as a text-book in normal schools. We are told that a course in general psychology is presupposed, though it is difficult to see why any intelligent student might not cope successfully with the text without such preliminary study. Each chapter is followed by a list of exercises, questions for discussion, and references for futher reading. "Constant emphasis has been thrown on the

physiological basis of the tendencies discussed, and Thorndike's classification of instincts, on the basis of the responses made, is adhered to throughout. The greatest space is devoted to a descriptive study of children as differentiated from adults." The chapter headings reveal something of the scope of the work. They are: The source of original nature; characteristics of original nature; tendencies resulting in action, non-social instincts; social instincts; tendencies accompanied by affective states; attention; sense perception; memory; imagination; thinking; habit and learning; play; moral and religious development; physical development of the child; a cross-section of child life at five and at eleven; exceptional children; and methods used in psychology. It will be seen that the authors follow fairly closely the rubrics of traditional psychology, yet examination of the text shows a wide divergence from traditional treatment. Throughout pains are taken to select from the rich literature of the topics treated those data that pertain specifically to the development of childhood. It is a distinct advance over any child psychology that we have seen.

PAUL POPENOE AND ROSWELL HILL JOHNSON. *Applied Eugenics.* New York: The Macmillan Company, 1918. Pp. xii, 459. $2.10.

This is a fascinating book on a subject of increasing social importance. The authors, one the editor of the *Journal of Heredity*, the other professor of sociology in the University of Pittsburgh, are admirably equipped to give a presentation of the subject both scientific and interesting. Their success has been marked. It is safe to predict that the book will speedily be recognized as the chief work on applied eugenics in any language. May it be widely read in order to counteract the many absurd beliefs that are held by otherwise intelligent people! One of the most interesting portions of the book is the chapter on The Eugenic Aspect of some Specific Reforms, including taxation, back to the farm, democracy, socialism, child labor, compulsory education, vocational guidance, minimum wage, mothers' pensions, housing, feminism, old age pensions, sex hygiene movement, trades unionism, prohibition, and pedagogical celibacy. Other important chapters are those on religion and eugenics, and eugenics and euthenics. Almost every chapter has something of significance for the student of education.

MABEL POWERS. *Stories the Iroquois Tell Their Children.* Cincinnati: The American Book Company, 1918. Pp. 216.

Nothing is more indicative of the temperament of a people than the stories which it tells to its children. The stories in this volume have been gathered at first hand from the surviving Indians in New York State, and their fidelity is attested by the chiefs of the tribes. They are interesting folk tales which gain an added attractiveness from the numerous illustrations, many of which are in color. The book is well adapted for supplementary reading in the intermediate grades.

KENNETH RICHMOND. *The Permanent Values in Education.* New York: E. P. Dutton and Company, 1917. Pp. xxiii, 136. $1.25.

This little book consists of a series of excursions into various fields of the history of education to discover and bring forth the lessons which educational movements of other days have for our own time Examination of the educational systems of the Jews, the Greeks, the Romans, the Renaissance, and the theories of Milton, Comenius, Locke, Rousseau, Pestalozzi, Herbart and Froebel, reveals certain funda-

mental principles which are valid for all time. The chief of these is that education must aim to make every man and every woman of a nation able to enjoy the entire fulness of life which the nation offers.

Schoolmen's Week Proceedings at the University of Pennsylvania. University of Pennsylvania Bulletin, Series 18, No. 5, Part 3, 1918. Pp. 362.

This volume contains many papers that are of interest to the scientific student of education. Among these are three addresses by Professor Charles H. Judd, of the University of Chicago, on "Educational Measurements—Their Benefits to Teachers and to Laymen," "Measurement of Reading," and "Measurement and the Promotion of Pupils;" "Report on Co-operative Work in Educational Measurements," by Harlan Updegraph and LeRoy A. King; "The Use of Educational Measurements in Formulating Changes in the Course of Study," by J. N. Adee; "Measuring the Fundamental Operations of Arithmetic in the Scranton Schools", by Thomas Francis; "The Courtis Tests in Arithmetic in a Group of Philadelphia Schools," by Philip A. Boyer; "A Study of Individual Pupils with the Aid of Courtis Tests in Arithmetic," by J. M. Fisher; "The Use of the Courtis Test in Silent Reading," by Katherine E. Moran; and "A Study in the Use of the Courtis Arithmetic Tests in Lock Haven," by C. W. Hunt.

F. L. AND A. M. SKINNER. *Happy Tales for Story Time.* New York: The American Book Company, 1918. Pp. 180. $0.64.

These delightful stories will provide many happy hours for tired and restless children. There are groups of animal tales, grandmother's tales, tales of Christmas, folk tales and fables, and wonder tales. The last story in the book is the Blue Bird, adapted from Maeterlinck.

GEORGE D. STRAYER AND FRANK P. BACHMAN. *The Gary Public Schools, Organization and Administration.* New York: General Education Board, 1918. Pp. xix, 126. $0.15.

This is the second volume of the Gary Survey. The introduction gives an outline of the Gary Plan and this is followed by a chapter setting forth present day problems, such as modernizing the curriculum, increasing the efficiency of instruction, and financing the educational enterprise. There follow chapters on program plant, organization, use of plant, supervision and administration, and comparative cost. The general supervision is characterized as inadequate in amount and ineffective in its results. The principals are administrative officers who have little to do with the educational aspects of the schools. The administration has new and difficult problems to meet, and has not shown great efficiency in their solution. This seems to be one of the weakest points in the Gary regime.

C. W. SUTTON. *The Results of a Spelling Test.* Cleveland: Board of Education, Division of Reference and Research, 1918. Pp. 19.

Lists of words were selected from the Ayres Spelling Scale and given to each grade from III to VIII. Each list contained from 30 to 50 words, and was composed of different words for each grade. This makes comparison of grades impossible. There were 1,941,185 spellings obtained from 50,687 children. A valuable feature of the record is the percentage of correct spellings on each word for each grade. The results of two years' drill upon the Ayres list of 1000 words are to make the records of the different schools much more uniform, and to change somewhat the relative position of the words in the scale of correct spellings.